

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 14, 2006, 02:19:42 ; Search time 6853 Seconds
(without alignments)
4424.052 Million cell updates/sec

Title: US-10-620-914-45

Perfect score: 3463

Sequence: 1 MCGRDRPASTYTKNFSLE.....RVNYSFFYMARKGAKGN 648

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Zgapop 6.0 , Zgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-O=/abes/ABSWEB spool/US10620914/runat_13032006_102042_6770/app_query.fasta_1
-DB=EST -OPMT=flatap -SUFFIX=p2n.rst -MINMATCH=0_1 -LOOPECL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptco -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abes05h
-USER=US10620914 @CGN 1 1 5315 @runat_13032006_102042_6770 -NCPU=6 -ICPU=3
-NO MMAP -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1296	37.4	743	5	BU651274
2	1003	29.0	556	1	AV627635
3	847	24.5	475	1	AV644266
4	817	23.6	467	1	AV629528
5	812	23.4	532	1	AV628087
6	686	20.1	391	1	AV626756
7	687	19.8	454	3	BP093208

8	664	19.2	379	3	BP086130
9	643	18.6	508	1	AV641303
10	631	18.2	352	3	BP093875
11	630	18.2	511	1	AV626139
12	627	18.1	354	1	AV631627
13	588	17.0	478	1	AV642589
14	584	16.9	480	1	AV643824
15	556	16.1	466	1	AV628989
16	531	15.3	461	5	BQ824076
17	518	15.0	447	1	AV642307
18	466	13.5	262	1	AV630897
19	418	12.1	398	3	BP093467
20	417	12.0	250	3	BP089550
21	412	11.9	388	3	BP095133
22	401	11.6	377	3	BP097104
23	374	10.8	362	3	BP096956
24	368	10.6	366	3	BP098395
25	353.5	10.2	433	11	AW561229
26	350.5	10.1	838	11	CNS076PM
27	348	10.0	973	11	CNS07804
28	331	9.6	846	8	DR662733
29	325	9.4	287	1	AV640882
30	311	9.0	850	11	CNS06VGM
31	309	8.9	963	11	CNS071FB
32	298	8.6	416	6	CF556716
33	291	8.4	176	1	AV628222
34	252	7.3	800	11	CNS072J0
35	244	7.0	585	3	BI721528
36	233	6.7	320	2	BI189160
37	232.5	6.7	564	1	AJ273357
38	220.5	6.4	1060	10	CW929808
39	219	6.3	301	2	BF072578
40	217.5	6.3	915	11	CNS079G1
41	212.5	6.1	511	9	BZ782470
42	205.5	5.9	817	7	CO023856
43	205.5	5.9	889	7	CO024454
44	202	5.8	877	11	CNS06VXF
45	183	5.3	567	7	CO151218

ALIGNMENTS

RESULT 1

BU651274

LOCUS

DEFINITION

BU651274

1112092D11.y1 C. reinhardtii

(normalized), Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA

sequence.

BU651274

BU651274.1

GI:23363454

Chlamydomonas reinhardtii

EST.

Chlamydomonas reinhardtii

Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;

Chlamydomonadales; Chlamydomonadales; Chlamydomonas

1 (bases 1 to 743)

Grossman, A., Chang, C.-W., Davies, J., Harris, E., Hauser, C.,

Lefebvre, P., McDermott, J., Shrager, J., Silflow, C. and Stern, D.

Analyses of the Chlamydomonas reinhardtii Genome: A Model,

Unicellular System for Analyzing Gene Function and Regulation in

Vascular Plants. Project: 1112

Unpublished (2002)

Contact: Charles Hauser

DCMB Box 91000

Durham, NC 27708-1000

Tel: 919 613 8159

Fax: 919 613 8177

Email: chauser@duke.edu.

Location/Qualifiers

1..743

/organism="Chlamydomonas reinhardtii"

/mol_type="mRNA"

/strain="21gr (CC-1690 wild type mt+) & 6145c (CC-1691 wild type mt-)"
 /db_xref="taxon:3055"
 /clone_lib="C. reinhardtii CC-1690 (mt+), CC-1691 (mt-), Gamete (normalized), Lambda Zap II"
 /note="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2: XhoI; Gamete library was constructed by Hui Zhao, Min Lu, Jeffrey McDermott, William J. Snell and John Davies. Strain 21gr cells (CC-1690; mating type plus) and strain 6145c cells (CC-1691; mating type minus) that had been growing on a light-dark cycle (13:11 L/D) in R-medium (Sager and Granick) were separately transferred into nitrogen-free medium at 8 hours into the light period. PolyA mRNA was purified from each sample every 2 hours for the next 18 hours. The mRNA was pooled and used for cDNA synthesis. The cDNA was directionally cloned into lambda zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites. pBluescript II SK- plasmids were excised from the lambda ZAP clones by superinfection with ExAssist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al., (1996) Genome Research 6: 791-806."

ORIGIN

Alignment Scores:
 Pred. No.: 6 23e-134 Length: 743
 Score: 1296.00 Matches: 243
 Percent Similarity: 99.2% Conservative: 0
 Best Local Similarity: 99.2% Mismatches: 2
 Query Match: 37.4% Indels: 1
 DB: 5 Gaps: 0

US-10-620-914-45 (1-648) x BU651274 (1-743)

Qy 228 GluGlnAsnThrGlnGlySerIleProTyrValProTyrLeuArgAlaProTyrTyrVal 247
 Db 8 GAGCAGAACACCCAGGCTTCGATCCCTACGTGCGGTGGTGGCGGCCCTACTACGTG 67
 Qy 248 TrpIleGlyArgLeuProSerValGlyHisAlaLeuHisGluGluArgValGluArgPro 267
 Db 68 TGGATTGGCGCGCTGCCACGGTTGGCCACGCCCTGCACGAGGAGCGCGTGGAGCGCGC 127
 Qy 268 ProMetPheProThrPheLeuTyrThrGlnSerTyrGluAspProGluProAspMet 287
 Db 128 CCATGTTCCCGCCACCTTCCTGTACACGAGTCTGGGAGGACCCCGACCGCATG 187
 Qy 288 GluValMetGluLeuAsnProLysAspThrValLeuThrLeuThrSerGlyGlyCysAsn 307
 Db 188 GAGGTGATGGAGATCAACCCCAAGGACACGCGTGTGACCTGACTAGCGCGGTGCAAT 247
 Qy 308 AlaLeuAsnLeuLeuValGlnGlyAlaGlyGlnValValSerValAspCysAsnProAla 327
 Db 248 GCCTGAACCTGCTGGTGCAGGGGGCGCGCGAGGTGTGCGTGGAGCTGCAACCCCGCG 307
 Qy 328 GlnSerAlaLeuLeuGluLeuLysValAlaIleGlnGlnLeuGluPheGluAspVal 347
 Db 308 CAGTCGCGGCTTCGGAGCTGAAGAAGGTGCCATTTCAGCAGCTGGAGTTTGAGCAGTG 367
 Qy 348 TrpGlnLeuPheGlyGluGlyValHisProArgIleGluGluLeuTyrGluLysLeu 367
 Db 368 TGGCAGCTGTTCGGCGAGGGCGTGCACCGCGCATTTGAGGAGCTGTACGAGAAGAGCTG 427
 Qy 368 AlaProPheLeuSerGlnThrSerHisAsnPheTrpSerLysArgLeuTyrTyrPheGln 387
 Db 428 CGCGCCCTTCCTGTCGCAACACAGCACAACCTTCGTGTCCTCAAGCGCCCTGTGATCTCCAG 487
 Qy 388 HisGlyLeuTyrTyrGlnGlyGlyMetGlyLysLeuCysTrpValLeuGlnCysLeuAla 407
 Db 488 CACGCGCTGTACTACAGGGCGGATGGCAGAGCTGTGCTGGGTGCTGCATGCTGGCC 547
 Qy 408 ValValLeuGlyLeuGlyLysThrValLysArgLeuAlaAsnAlaProThrMetGluGlu 427
 Db 548 GTGGTGTGGGACTGGGCAAGACCGTCAAGCGGCTCGCCCAACCGGCCCAATGGAGGAG 607

Qy 428 GlnArgArgLeuTyrAspSerAsnMetLeuIleHisPheValLysAsnGlyProLysPro 447
 Db 608 CAGCGCGCTGTGGAGCAGCAACATGCTCATCTTCGTGAGAACGGGCCCAAGCCG 667
 Qy 448 LeuValTrpLeuPheValLysPheValSerLeuValLeuPheAsnLysAlaValLeuTyr 467
 Db 668 CTGGTGTGGCTGTTCGTCAAGTTTCGTGAGCTGCTCTTCAACAG-GCCGTGCTGTGG 726
 Qy 468 PheGlyGlyGlyVal 472
 Db 727 TTCNCGCGCGCGGTG 741
 RESULT 2
 AV627635 556 bp mRNA linear EST 15-DEC-2000
 LOCUS AV627635 Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas
 DEFINITION reinhardtii cDNA clone LCL029g11_r 5', mRNA sequence.
 AV627635
 VERSION AV627635.1 GI:10790269
 KEYWORDS EST.
 SOURCE Chlamydomonas reinhardtii
 ORGANISM Chlamydomonas reinhardtii
 Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;
 Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.
 REFERENCE 1 (bases 1 to 556)
 AUTHORS Asamizu, E., Miura, K., Kucho, K., Inoue, Y., Fukuzawa, H., Ohya, K.,
 Nakamura, Y. and Tabata, S.
 TITLE Generation of expressed sequence tags from low-CO2 and high-CO2
 adapted cells of Chlamydomonas reinhardtii
 JOURNAL DNA Res. 7 (5), 305-307 (2000)
 PUBMED 11089912
 COMMENT Contact: Erika Asamizu
 The First Laboratory for Plant Gene Research
 Kazusa DNA Research Institute
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
 Email: asamizuekazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
 FEATURES
 source
 1. 556
 /organism="Chlamydomonas reinhardtii"
 /mol_type="mRNA"
 /strain="C9"
 /db_xref="taxon:3055"
 /clone_lib="Chlamydomonas reinhardtii 5% to 0.04% CO2"
 /note="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2:
 XhoI; The cDNA library was constructed from cells cultured
 in a carbon stress acclimatized condition in which carbon
 dioxide concentration in the bubbling gas was changed from
 5% to 0.04%"
 ORIGIN
 Alignment Scores:
 Pred. No.: 2 67e-101 Length: 556
 Score: 1003.00 Matches: 184
 Percent Similarity: 100.0% Conservative: 0
 Best Local Similarity: 100.0% Mismatches: 0
 Query Match: 29.0% Indels: 0
 DB: 1 Gaps: 0
 US-10-620-914-45 (1-648) x AV627635 (1-556)
 Qy 96 LysPheLysSerIleTyrValValAspLeuCysHisSerLeuCysGluValAlaLysLys 115
 Db 3 AAGTTCAAGTCCATCTACGTGGTGGACCTGTGTCACCTGCTGCGAGGTGGCCAAAGAG 62
 Qy 116 LysAlaLysAlaLysGlyTrpLysAsnValGlnValGluAlaAspAlaCysGlnPhe 135
 Db 63 AAGCGAAGCCCAAGGCGCTGGAAGATGTCCAGTGTGGAGCGCGCTTGCCTTTCCTT 122
 Qy 136 AlaProProGluGlyThrAlaThrLeuIleThrPheSerTyrSerLeuThrMetIlePro 155
 Db 123 GCGCCCTTGGAGGACCGCGCGCTCATCACCTTCTCTACTGCTCAGCATGATTTCCA 182

ORIGIN

Alignment Scores: 1.41e-80 Length: 467
Pred. No.: 817.00 Matches: 154
Score: 817.00 Conservative: 150
Percent Similarity: 99.4% Mismatches: 1
Best Local Similarity: 99.4% Indels: 0
Query Match: 23.6% Gaps: 0
DB: 1

US-10-620-914-45 (1-648) x AV629528 (1-467)

QY 60 AlaAlaPheAlaAlaArgLeuAlaGluArgSerAsnLeuIleTrpValAspLeuGlyGly 79
DB 1 GCTGCAGTTGCTGCCCGCTGGCGAGCGCTCGAACCTCATCTGGGTGACCTGGGTGGT 60
QY 80 GlyThrGlyGluAsnValAspMetMetAlaAspTyrIleAspLeuAlaLysPheLysSer 99
DB 61 GGCACCTGGGAGAAATTCGATATGCTGCTGATTACATCGACCTGGCGAAGTTCAAGTCC 120
QY 100 IleTyrValValAspLeuCysHisSerLeuCysGluValAlaLysLysLysAlaLysAla 119
DB 121 ATCTACGTGTGTCACCTGTGCCACTGCTGTGAGGTGGCCGAAGAGGCGAAGGCC 180
QY 120 LysGlyTyrLysAsnValGlnValValGluAlaAspAlaCysGlnPheAlaProProGlu 139
DB 181 AAGGCTGGGAAGATGTCAGGTCTGTGAGCGCCGACGCTTGCCAAATTTGGCCCCCTGAG 240
QY 140 GlyThrAlaThrIleuIleThrPheSerTyrSerLeuThrMetIleProPheHisAsn 159
DB 241 GGCACCGCGAGCTCATCATCTTCTCTACTCGCTCAGCATGATTCACCGCTTCCCAAC 300
QY 160 ValIleAspGlnAlaCysSerTyrLeuSerGlnAspGlyLeuValGlyValAlaAspPhe 179
DB 301 GTCATGACGAGGCTTCTCGTACCTGTCCAGACGCGCTGTGGCGGTTCGCGACTTC 360
QY 180 TyrValSerGlyLysTyrAspLeuProLeuArgGlnMetProTyrSerArgArgPhePhe 199
DB 361 TACGTAGCGGCAAGTACGACCTGCCCTCGCCAGATGCCCTGGTGGCGGTTCTTCTTC 420
QY 200 TrpArgSerIlePheAspIleAspAsnIleAspIleGlyProGlu 214
DB 421 TGGCGATCGATCTTCGATCGACATCGAACATTCGATCGCGCCCCGAG 465

RESULT 5
LOCUS AV628087 532 bp mRNA linear EST 15-DEC-2000
DEFINITION AV628087 Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas
reinhardtii cDNA clone LCL036e03_r 5', mRNA sequence.

ACCESSION AV628087

VERSION AV628087.1 GI:10790721

KEYWORDS EST.

SOURCE Chlamydomonas reinhardtii

ORGANISM Chlamydomonas reinhardtii

Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;

Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.

1 (bases 1 to 532)

Asamizu,E., Miura,K., Kucho,K., Inoue,Y., Fukuzawa,H., Ohyama,K.,

Nakamura,Y. and Tabata,S.

Generation of expressed sequence tags from low-CO2 and high-CO2

adapted cells of Chlamydomonas reinhardtii

DNA Res. 7 (5), 305-307 (2000)

PUBMED 11089912

COMMENT Contact: Erika Asamizu

The First Laboratory for Plant Gene Research

Kazusa DNA Research Institute

Yana 1532-3, Kisarazu, Chiba 292-0812, Japan

Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

Location/Qualifiers

FEATURES

source

1..532 /organism="Chlamydomonas reinhardtii"

/mol_type="mRNA"

/strain="C9"

/db_xref="taxon:3055"

/clone_lib="LCL036e03_r"

/note="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2:

/note="The cDNA library was constructed from cells cultured
in a carbon stress acclimatized condition in which carbon
dioxide concentration in the bubbling gas was changed from
5% to 0.04%"

ORIGIN

Alignment Scores: 6.32e-80 Length: 532
Pred. No.: 812.00 Matches: 158
Score: 812.00 Conservative: 1
Percent Similarity: 89.8% Mismatches: 0
Best Local Similarity: 89.3% Indels: 18
Query Match: 23.4% Gaps: 1
DB: 1

US-10-620-914-45 (1-648) x AV628087 (1-532)

QY 3 SerGlyArgAspGlyArgProAlaSerTyrThrLysLysAsnPheSerLeuGluLysLeu 22
DB 1 TCGGTCGTGACGCGCGCTGCGAGCTACACCAAGAGAACTTCTCCCTGGAGAGCTC 60
QY 23 LysLeuSerSerMetLysAspLeuThrValLeuArgHisMetTrpPheGlySerLys 42
DB 61 AAGCTCAGCAGCATGAAGGATGACCTGACCGTTCTCGCCATATGTGTTTCGGCAGCAAG 120
QY 43 LysGlyAspAspHisAlaAlaArgLeuGluSerPheTyrGlyProGlnAlaAlaPhe 62
DB 121 AAGGGCGATGATCACGCTGCTCGCTGGAGAGCTTCTACGGGCCCCCAGCGCGCTGCTAT 180
QY 63 -----AlaAla 64
DB 181 GATGCTTTCGGTCGGGTTCTCTCGGGTCGCGAGCCCATGCTCGCTGCGAGTTGCTGCC 240
QY 65 ArgLeuAlaGluArgSerAsnLeuIleTrpValAspLeuGlyGlyThrGlyGluAsn 84
DB 241 CGCCTGGCGAGCGCTCGAACCTCATCTGGGTTGACCTGGGTGGTGGCACTGGGGAGAA 300
QY 85 ValAspMetMetAlaAspTyrIleAspLeuAlaLysPheLysSerIleTyrValValAsp 104
DB 301 GTCGATATGATGGCTGATTACATGACCTGGCGAAGTTCAAGTCCATCTACGTGGTGCAC 360
QY 105 LeuCysHisSerLeuCysGluValAlaLysLysLysLysLysLysLysLysLysLys 124
DB 361 CTGTGCACTCGCTGTGCGAGGTGGCCAGAGAGCGGAGCGGAGGCTGGAGAAAT 420
QY 125 ValGlnValValGluAlaAspAlaCysGlnPheAlaProProGluGlyThrAlaThrLeu 144
DB 421 GTCCAGGTGCTGGAGGCGGACGCTTGCCAATTTGCGCCCCCTGAGGCGACCGGACGCTC 480
QY 145 IleThrPheSerTyrSerLeuThrMetIleProPheHisValIle 161
DB 481 ATCACCTTCTCTACTCGCTCAGCATGATTCACCGCTTCCACACGTCATC 531

RESULT 6

LOCUS AV626756

DEFINITION AV626756 Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas

reinhardtii cDNA clone LCL014h09_r 5', mRNA sequence.

ACCESSION AV626756

VERSION AV626756.1 GI:10789036

KEYWORDS EST.

SOURCE Chlamydomonas reinhardtii

ORGANISM Chlamydomonas reinhardtii

Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;

Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.

1 (bases 1 to 391)

Asamizu,E., Miura,K., Kucho,K., Inoue,Y., Fukuzawa,H., Ohyama,K.,

Nakamura,Y. and Tabata,S.

Generation of expressed sequence tags from low-CO2 and high-CO2

adapted cells of Chlamydomonas reinhardtii

JOURNAL DNA Res. 7 (5), 305-307 (2000)
 PUBMED 11089912
 COMMENT Contact: Erika Asamizu
 The First Laboratory for Plant Gene Research
 Kazusa DNA Research Institute
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
 Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES
 source
 1. .391
 /organism="Chlamydomonas reinhardtii"
 /mol_type="mRNA"
 /strain="C9"
 /db_xref="taxon:3055"
 /clone_lib="ICL014h09_r"
 /clone="1532-3, Kisarazu, Chiba 292-0812, Japan
 /note="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2:
 XhoI; The cDNA library was constructed from cells cultured
 in a carbon stress acclimatized condition in which carbon
 dioxide concentration in the bubbling gas was changed from
 5% to 0.04%"

ORIGIN
 Alignment Scores:
 Pred. No.: 3 85e-67 Length: 391
 Score: 696.00 Matches: 130
 Percent Similarity: 100.0% Conservative: 0
 Best Local Similarity: 100.0% Mismatches: 0
 Query Match: 20.1% Indels: 0
 DB: 1 Gaps: 0

US-10-620-914-45 (1-648) x AV626756 (1-391)

QY 77 LeuGlyGlyThrGlyGluAsnValAspMetAlaAspTyrIleAspLeuAlaLys 96
 DB 1 CTGGTGGTGGCACTGGGAGATGCGATATGATGCTGATTCATCGACCTGGCGAAG 60

QY 97 PheLysSerIleTyrValValAspLeuCysHisSerLeuCysGluValAlaLysLys 116
 DB 61 TTCAGTCCATCTACGACCGGCTGCGACCTGTCCTGCTGCGAGGTGGCGAAGAAG 120

QY 117 AlaLysAlaLysGlyTyrLysAsnValGlnValAlaAspAlaCysGlnPheAla 136
 DB 121 GCGAGCGCCAGGCGCTGGAGATGTCAGGTGTCGAGCGCCGACCTTGCCTTGGC 180

QY 137 ProProGluGlyThrAlaThrLeuIleThrPheSerTyrSerLeuThrMetIleProPro 156
 DB 181 CCCCTGAGGCGACCGGAGCTCATCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 240

QY 157 PheHisAsnValIleAspGlnAlaCysSerTyrLeuSerGlnAspGlyLeuValGlyVal 176
 DB 241 TTCACACAGCTCATCGACCGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300

QY 177 AlaAspPheTyrValSerGlyLysTyrAspLeuProLeuArgGlnMetProTyrSerArg 196
 DB 301 GCGGACTTCTACGTAGCGGCAAGTACGACCTGCTGCTGCTGCTGCTGCTGCTGCTG 360

QY 197 ArgPhePheTyrArgSerIlePheAspIle 206
 DB 361 CGTTCTTCTGGCGATCGATCTTCGACATC 390

RESULT 7
 BP093208
 LOCUS BP093208 Chlamydomonas reinhardtii cDNA clone MXL005a04_r 5', mRNA sequence.
 DEFINITION Chlamydomonas reinhardtii cDNA clone MXL005a04_r 5', mRNA sequence.
 ACCESSION BP093208
 VERSION BP093208.1 GI:49465295
 KEYWORDS EST.
 SOURCE Chlamydomonas reinhardtii
 ORGANISM Chlamydomonas reinhardtii
 Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;
 Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.
 1 (bases 1 to 454)

AUTHORS Asamizu, E., Nakamura, Y., Miura, K., Fukuzawa, H., Fujiwara, S.,
 Hirono, M., Iwamoto, K., Matsuda, Y., Minagawa, J., Shimogawara, K.,
 Takahashi, Y. and Tabata, S.
 TITLE Establishment of Publicly Available cDNA Material and Information
 Resource of Chlamydomonas reinhardtii (Chlorophyta), to Facilitate
 Gene Function Analysis
 JOURNAL Phycologia (2004) in press
 COMMENT The First Laboratory for Plant Gene Research
 Kazusa DNA Research Institute
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
 Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES
 source
 1. .454
 /organism="Chlamydomonas reinhardtii"
 /mol_type="mRNA"
 /strain="C9"
 /db_xref="taxon:3055"
 /clone_lib="MXL005a04_r"
 /clone="MXL005a04_r"
 /note="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2:
 XhoI; The cDNA library was made from a mixture of cells
 grown under various conditions"

ORIGIN
 Alignment Scores:
 Pred. No.: 4 99e-66 Length: 454
 Score: 687.00 Matches: 132
 Percent Similarity: 88.1% Conservative: 1
 Best Local Similarity: 87.4% Mismatches: 0
 Query Match: 19.8% Indels: 18
 DB: 3 Gaps: 1

US-10-620-914-45 (1-648) x BP093208 (1-454)

QY 36 HisMetTyrPheGlySerLysGlyAspAspHisAlaAlaArgLeuGluSerPheTyr 55
 DB 1 CATATGTGTTGGCGAGCAAGAGGCGATGATGATGCTGCTGCTGCTGCTGCTGCTGCT 60

QY 56 GlyProGlnAlaAlaAlaPhe----- 62
 DB 61 GGGCCCCAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120

QY 63 -----AlaAlaArgLeuAlaGluArgSerAsnLeuIleTyrValAspLeu 77
 DB 121 ATGCTCGTGCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180

QY 78 GlyGlyGlyThrGlyGluAsnValAspMetMetAlaAspTyrIleAspLeuAlaLysPhe 97
 DB 181 GGTGGTGGCACTGGGAGAAATGCGATATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240

QY 98 LysSerIleTyrValValAspLeuCysHisSerLeuCysGluValAlaLysLysAla 117
 DB 241 AAGTCCATCTACGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300

QY 118 LysAlaLysGlyTyrLysAsnValGlnValAlaAlaAspAlaCysGlnPheAlaPro 137
 DB 301 AAGCCNAGGCGTGGAGAAATGTCAGTCTGCGAGGCGCGAGCGCTTGCCTTGGCGCC 360

QY 138 ProGluGlyThrAlaThrLeuIleThrPheSerTyrSerLeuThrMetIleProPhe 157
 DB 361 CCTGAGGCGACGCGACGCTCATCCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 420

QY 158 HisAsnValIleAspGlnAlaCysSerTyrLeu 168
 DB 421 CACAACGTCATCGACCGGCTTGTCTGTACCTG 453

RESULT 8
 BP086130
 LOCUS BP086130 Chlamydomonas reinhardtii cDNA clone MX002e03_r 5', mRNA sequence.
 DEFINITION Chlamydomonas reinhardtii cDNA clone MX002e03_r 5', mRNA sequence.

BP086130
 VERSION BP086130.1 GI:49458217
 EST.
 Chlamydomonas reinhardtii
 Chlamydomonas reinhardtii
 Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;
 Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.
 1 (bases 1 to 379)
 Asamizu,E., Nakamura,Y., Miura,K., Fukuzawa,H., Fujiwara,S.,
 Hirono,M., Iwanoto,K., Matsuda,Y., Minagawa,J., Shimogawara,K.,
 Takahashi,Y. and Tabata,S.
 Establishment of Publicly Available cDNA Material and Information
 Resource of Chlamydomonas reinhardtii (Chlorophyta), to Facilitate
 Gene Function Analysis
 Phycologia (2004) In press
 Contact: Erika Asamizu
 The First Laboratory for Plant Gene Research
 Kazusa DNA Research Institute
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
 Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES
 source
 1..379
 /organism="Chlamydomonas reinhardtii"
 /mol_type="mRNA"
 /strain="C9"
 /db_xref="taxon:3055"
 /clone="WX002e03_r"
 /clone_lib="Chlamydomonas reinhardtii C9 various
 conditions"
 /notes="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2:
 XhoI; The cDNA library was made from a mixture of cells
 grown under various conditions"

ORIGIN

Alignment Scores:
 Pred. No.: 1,41e-63 Length: 379
 Score: 664.00 Matches: 126
 Percent Similarity: 100.0% Conservative: 0
 Best Local Similarity: 100.0% Mismatches: 0
 Query Match: 19.2% Indels: 0
 DB: 3 Gaps: 0

US-10-620-914-45 (1-648) x BP086130 (1-379)

Qy 484 GlyIleProTleGluAnTYrIleAlaArgThrMetAspGlyValAlaGluAenSerHis 503
 Db 2 GGCATCCCCATTGAGAACTACATCGCGCGCACCATGGACCGCGTGGCGAGAACTCGCAC 61
 Qy 504 ValArgLysGlnAnTYrPheTYrTYrAsnCysLeuThrGlyLysPheLeuArgAspAsn 523
 Db 62 GTGGCGAAGCAGAACTACTTCTACTACAACTGCTCACCGGCAAGTTCCTCGCGGACAAAC 121
 Qy 524 CysProThrTYrLeuArgGluAlaAlaPheAlaThrLeuLysSerGlyValValAspAsn 543
 Db 122 TGCCCCACCTACCTGCGCGAGGCGGCTTCGCCACCCTCAAGAGTGGCGTGGTGGCAAC 181
 Qy 544 LeuThrValSerThrAnPhePheMetGluLeuLysAlaArgThrTYrThrLysVal 563
 Db 182 CTGACCGCTCTCCACCACTTCTTCATGGAGAGCTCAAGCGCGCACCTACACCAAGGTG 241
 Qy 564 IleLeuMetAspHisValAspTrpLeuAspMetProValAlaAenGluLeuAlaGluCys 583
 Db 242 ATTCTGATGACCAACGTTGGACTGGCTGGATATGCCGTGGCCCAACGAGCTGGCCGAGTGC 301
 Qy 584 LeuAlaLysGlnValAlaProGlyGlyIleValIleTrpArgSerAlaSerLeuSerPro 603
 Db 302 CTGGCCCAAGCAGGTGGCGCGGCGGATCGTCATCTGGCGCTCCGCTCCCTCAGCCCG 361
 Qy 604 ProTYrAlaGluLeuIle 609
 Db 362 CCCTACCGCCGAGCTGATC 379

RESULT 9

AV641303
 LOCUS
 DEFINITION AV641303 Chlamydomonas reinhardtii 5' CO2 linear EST 15-DEC-2000
 cDNA clone HCL031d02_r 5', mRNA sequence.
 AV641303
 VERSION AV641303.1 GI:10784631
 EST.
 Chlamydomonas reinhardtii
 Chlamydomonas reinhardtii
 Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;
 Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.
 1 (bases 1 to 508)
 Asamizu,E., Miura,K., Kucho,K., Inoue,Y., Fukuzawa,H., Ohyama,K.,
 Nakamura,Y. and Tabata,S.
 Generation of expressed sequence tags from low-CO2 and high-CO2
 adapted cells of Chlamydomonas reinhardtii
 DNA Res. 7 (5), 305-307 (2000)
 Contact: Erika Asamizu
 The First Laboratory for Plant Gene Research
 Kazusa DNA Research Institute
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
 Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES
 source
 1..508
 /organism="Chlamydomonas reinhardtii"
 /mol_type="mRNA"
 /strain="C9"
 /db_xref="taxon:3055"
 /clone="HCL031d02_r"
 /clone_lib="Chlamydomonas reinhardtii 5' CO2"
 /notes="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2:
 XhoI; The cDNA library was constructed from cells cultured
 in a medium with bubbling air containing 5% carbon
 dioxide"

ORIGIN

Alignment Scores:
 Pred. No.: 5.1e-61 Length: 508
 Score: 643.00 Matches: 126
 Percent Similarity: 87.6% Conservative: 1
 Best Local Similarity: 86.9% Mismatches: 0
 Query Match: 18.6% Indels: 18
 DB: 1 Gaps: 1

US-10-620-914-45 (1-648) x AV641303 (1-508)

Qy 1 MetGlySerGlyArgAspGlyArgProAlaSerTYrThrLysLysAsnPheSerLeuGlu 20
 Db 72 ATGGGTTCGGTCTGTCAGCGCGCTCGAGCTACACCAAGAAGAACTTCTCCCTGGAG 131
 Qy 21 LysLeuLysLeuSerSerMetLysAspAspLeuThrValLeuArgHisMetTrpPheGly 40
 Db 132 AAGCTCAAGCTCAGCAGCATGAAGGATGACCTGACCGCTTCTGGCCCATATGTGTTCCGC 191
 Qy 41 SerLysLysGlyAspAspHisAlaAlaArgLeuGluSerPheTYrGlyProGlnAlaAla 60
 Db 192 AGCAAGAGGCGCATGATCACGCTGCTCGCTGGAGAGCTTCTACGGGCCCCAGCCGCT 251
 Qy 61 AlaPhe----- 62
 Db 252 GCCTATGATGCTTTCGGTTCGGTTCCTCTGGGGTTCGAGGCCCATGCTCGTGCAGTT 311
 Qy 63 AlalaArgLeuAlaGluArgSerAsnLeuIleTrpValAspLeuGlyGlyThrGly 82
 Db 312 GCTGCCCGCTGGCCGAGCGCTCGAACCTCATCTGGGTTGACCTGGGTGGTGGCAGCTGGG 371
 Qy 83 GluAsnValAspMetMetAlaAspTYrIleAspLeuAlaLysPheLysSerIleTYrVal 102
 Db 372 GAGATGTCGATATGATGGCTGATTACATCGACCTGCGGAGTTCAAGTCCATCTACGTG 431
 Qy 103 ValAspLeuCysHisSerLeuGluValAlaLysLysLysAlaLysGlyTrp 122

```

Db      432  GTCACCTGTGCGACTCGCTGTGCGAGTGTCGCCAAGAGGCGCAAGGCGCAAGGGCTGG 491
Qy      123  LysAenValGlnVal 127
Db      492  AAGAATGTCCAGGTC 506

RESULT 10
BP093875
LOCUS      352 bp mRNA linear EST 30-JUN-2004
DEFINITION Chlamydomonas reinhardtii C9 various conditions
Chlamydomonas reinhardtii cDNA clone MXL015a05_r 5', mRNA sequence.
ACCESSION BP093875
VERSION BP093875.1 GI:49465962
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii
ORGANISM Chlamydomonas reinhardtii
REFERENCE 1 (bases 1 to 352)
AUTHORS Asamizu, E., Nakamura, Y., Miura, K., Fukuzawa, H., Fujiwara, S.,
Hirono, M., Iwamoto, K., Matsuda, Y., Minagawa, J., Shimogawara, K.,
Takahashi, Y. and Tabata, S.
TITLE Establishment of Publicly Available cDNA Material and Information
Resource of Chlamydomonas reinhardtii (Chlorophyta), to Facilitate
Gene Function Analysis
JOURNAL Phycologia (2004) In press
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES
source
1. 352
/organism="Chlamydomonas reinhardtii"
/mol_type="mRNA"
/strain="C9"
/db_xref="taxon:3055"
/clone="MXL015a05_r"
/clone_lib="Chlamydomonas reinhardtii C9 various
conditions"
/note="vector: pBluescriptII SK-; Site 1: EcoRI; Site 2:
XhoI; The cDNA library was made from a mixture of cells
grown under various conditions"

ORIGIN
Alignment Scores:
Pred. No.: 6,26e-60 Length: 352
Score: 631.00 Matches: 117
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 18.2% Indels: 0
DB: 3 Gaps: 0

US-10-620-914-45 (1-648) x BP093875 (1-352)

Qy      420  AlaAenAlaProThrMetGluGlnArgLeuTrpAspSerAenMetLeuHis 439
Db      2  GCCAACGCGCCCAATATGAGGAGGACGCGCGTGTGGGACAGCAACATGCTCATCCAC 61
Qy      440  PheValLysAnGlyProLysProLeuValTrpLeuPheValLysPheValSerLeuVal 459
Db      62  TTCGTGAAGAACGGGCCCAAGCCGCTGTGTGGCTGTTCGTAAGTTCTGTGAGCCTGGTG 121
Qy      460  LeuPheAenLysAlaValLeuTrpPheGlyGlyValProGlyLysGlnTrpAlaLeu 479
Db      122  CTCTTCAACAGCCGCGTGTGTTCGGCGCGCGCTGCGCGGCAAGCAAGTACGCGCTG 181
Qy      480  IleLysAlaAspGlyIleProLysGlnAsnTrpIleAlaArgThrMetAspGlyValAla 499
Db      182  ATCAAGCGGACGGCATCCCATTTGAGAACTACATCGCGCGCACCATGAGACGGCGTGGCG 241
Qy      500  GluAenSerHisValArgLysGlnAsnTrpPheTrpTrpAenCysLeuThrGlyLysPhe 519

```

```

Db      242  GAGAACTGCGACGTGCGCAAGCAGAACTACTTCTACTACAACTGCCTCACCAGCAAGTTC 301
Qy      520  LeuArgAspAenCysProThrTrpLeuArgGluAlaAalaPheAlaThrLeu 536
Db      302  CTGGCGGACAACTGCCCACTTACCTGCGGAGGCGGCTTCGCCACCTC 352

RESULT 11
AV626139
LOCUS      511 bp mRNA linear EST 15-DEC-2000
DEFINITION Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas
reinhardtii cDNA clone LCL003a03_r 5', mRNA sequence.
ACCESSION AV626139
VERSION AV626139.1 GI:10788419
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii
ORGANISM Chlamydomonas reinhardtii
REFERENCE 1 (bases 1 to 511)
AUTHORS Asamizu, E., Miura, K., Kucho, K., Inoue, Y., Fukuzawa, H., Ohyama, K.,
Nakamura, Y. and Tabata, S.
TITLE Generation of expressed sequence tags from low-CO2 and high-CO2
adapted cells of Chlamydomonas reinhardtii
JOURNAL DNA Res. 7 (5), 305-307 (2000)
PUBMED 11089912
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES
source
1. 511
/organism="Chlamydomonas reinhardtii"
/mol_type="mRNA"
/strain="C9"
/db_xref="taxon:3055"
/clone="LCL003a03_r"
/clone_lib="Chlamydomonas reinhardtii 5% to 0.04% CO2"
/note="vector: pBluescriptII SK-; Site 1: EcoRI; Site 2:
XhoI; The cDNA library was constructed from cells cultured
in a carbon stress acclimatized condition in which carbon
dioxide concentration in the bubbling gas was changed from
5% to 0.04%"

ORIGIN
Alignment Scores:
Pred. No.: 1,47e-59 Length: 511
Score: 630.00 Matches: 123
Percent Similarity: 87.3% Conservative: 1
Best Local Similarity: 86.6% Mismatches: 0
Query Match: 18.2% Indels: 18
DB: 1 Gaps: 1

US-10-620-914-45 (1-648) x AV626139 (1-511)

Qy      1  MetGlySerGlyArgAspGlyArgProAlaSerTyrThrLysLysAenPheSerLeuGlu 20
Db      84  ATGGGTGCGGTGCTGACGCGCGCTGCGAGCTACACCAAGAGAACTTCTCCCTGGAG 143
Qy      21  LysLeuLysLeuSerSerSerMetLysAspAspLeuThrValLeuArgHisMetTrpPheGly 40
Db      144  AAGCTCAAGCTCAGCAGCATGAAGATGACCTGACCGTTCTGCGCCATATATGTGTTCCGC 203
Qy      41  SerLysLysGlyAspAspHisAlaAlaArgLeuGluSerPheTyrGlyProGlnAlaAla 60
Db      204  AGCAAGAGGCGCATGATCAGCTGCTGCGCTGAGAGCTTCTACGGGCCCCAGGCGCT 263
Qy      61  AlaPhe-----
Db      264  GCCTATGATGCTTTCGGTTCGGTTCCTTCGGGTTCGAGGCCCATGCTCGCTGCAGTT 323
Qy      63  AlaAlaArgLeuAlaGluArgSerAenLeuIleTrpValAspLeuGlyGlyThrGly 82

```

Db 324 GCTGCCCGCTGGCCGAGCGCTCGAACCTCATCTGGGTTGACCTGGGTGGTGGCACTGGG 383
 Qy 83 GluAsnValAspMetAlaAspTyrIleAspLeuAlaLysPheLysSerIleTyrVal 102
 Db 384 GAGAATGTCATGATGATGGCTGATTCATGACCTGGCGAAGTTCAGTCCATCTACGTG 443
 Qy 103 ValAspLeuCyHisSerLeuCyGluValAlaLysLysLysAlaLysGlyTyr 122
 Db 444 GTCGACCTGTGCCACTCGCTGTGGAGGTGGCCCAAGAGGCGCAAGGGCTGG 503
 Qy 123 LysAsn 124
 Db 504 AAGAAT 509

RESULT 12

AV631627 354 bp mRNA linear EST 15-DEC-2000
 LOCUS AV631627 Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas
 DEFINITION reinhardtii cDNA clone LCL097c01_r 5', mRNA sequence.
 ACCESSION AV631627
 VERSION AV631627.1 GI:10794261
 KEYWORDS EST.
 SOURCE Chlamydomonas reinhardtii
 ORGANISM Chlamydomonas reinhardtii
 Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;
 Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.
 REFERENCE 1 (bases 1 to 354)
 AUTHORS Asamizu,E., Miura,K., Kucho,K., Inoue,Y., Fukuzawa,H., Ohyama,K.,
 Nakamura,Y. and Tabata,S.
 TITLE Generation of expressed sequence tags from low-CO2 and high-CO2
 adapted cells of Chlamydomonas reinhardtii
 JOURNAL DNA Res. 7 (5), 305-307 (2000)
 PUBMED 11089912
 COMMENT Contact: Erika Asamizu
 The First Laboratory for Plant Gene Research
 Kazusa DNA Research Institute
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
 Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES

source
 1..354
 /organism="Chlamydomonas reinhardtii"
 /mol_type="mRNA"
 /strain="C9"
 /db_xref="taxon:3055"
 /clone="LCL097c01_r"
 /clone_lib="Chlamydomonas reinhardtii 5% to 0.04% CO2"
 /note="vector: pBluescriptII SK-; Site 1: EcoRI; Site 2:
 XhoI; The cDNA library was constructed from cells cultured
 in a carbon stress acclimatized condition in which carbon
 dioxide concentration in the bubbling gas was changed from
 5% to 0.04%"

ORIGIN

Alignment Scores:
 Pred. No.: 1.77e-59 Length: 354
 Score: 627.00 Matches: 117
 Percent Similarity: 100.0% Conservative: 0
 Best Local Similarity: 100.0% Mismatches: 0
 Query Match: 18.1% Indels: 0
 DB: 1 Gaps: 0

US-10-620-914-45 (1-648) x AV631627 (1-354)

Qy 111 GluValAlaLysLysLysAlaLysGlyTyrLysAsnValGlnValGluAla 130
 Db 3 GAGGTGGCCAGAGAAGGCGAAGCGGCTGGGAAGATGTCCAGGCTGGTGGAGGCC 62
 Qy 131 AspAlaCysGlnPheAlaProGluGlyThrAlaThrLeuIleThrPheSerTyrSer 150
 Db 63 GACGCTTGCCAAATTTGGCCCCCTTGAGGGCACCGGAGCTCATCCTTCTCTACTCG 122
 Qy 151 LeuThrMetIleProPropheHisAsnValIleAspGlnAlaCysSerTyrLeuSerGln 170

Db 123 CTCACGATGATTCACCCGTTCACACGTCATCGACAGGCTTGCTGACCTGTCCCAA 182
 Qy 171 AspGlyLeuValGlyValAlaAspPheTyrValSerGlyLysTyrAspLeuProLeuArg 190
 Db 183 GACGGCTGTGGGCGTTTCCGCACTTCTACGTGAGCGGCAAGTACGACCTGCCCTGGC 242
 Qy 191 GlnMetProTyrSerArgPheTyrArgSerIlePheAspIleAspAsnIleAsp 210
 Db 243 CAGATGCCCTGGTGGCGCGTTTCTTCGCGATCGATCTTCGACATCGACAACATTGAC 302
 Qy 211 IleGlyProGluArgAlaTyrIleGluGlnLysLeuGluArgValTyr 227
 Db 303 ATCGGCCCGAGCGCGCGCTTACCTGGAGCAGAAGCTGGAGCGCGTGTGG 353

RESULT 13

AV642589 478 bp mRNA linear EST 15-DEC-2000
 LOCUS AV642589 Chlamydomonas reinhardtii 5% CO2 Chlamydomonas reinhardtii
 DEFINITION cDNA clone HCL054c11_r 5', mRNA sequence.
 ACCESSION AV642589
 VERSION AV642589.1 GI:10785917
 KEYWORDS EST.
 SOURCE Chlamydomonas reinhardtii
 ORGANISM Chlamydomonas reinhardtii
 Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;
 Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.
 REFERENCE 1 (bases 1 to 478)
 AUTHORS Asamizu,E., Miura,K., Kucho,K., Inoue,Y., Fukuzawa,H., Ohyama,K.,
 Nakamura,Y. and Tabata,S.
 TITLE Generation of expressed sequence tags from low-CO2 and high-CO2
 adapted cells of Chlamydomonas reinhardtii
 JOURNAL DNA Res. 7 (5), 305-307 (2000)
 PUBMED 11089912
 COMMENT Contact: Erika Asamizu
 The First Laboratory for Plant Gene Research
 Kazusa DNA Research Institute
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
 Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES

source
 1..478
 /organism="Chlamydomonas reinhardtii"
 /mol_type="mRNA"
 /strain="C9"
 /db_xref="taxon:3055"
 /clone="HCL054c11_r"
 /clone_lib="Chlamydomonas reinhardtii 5% CO2"
 /note="vector: pBluescriptII SK-; Site 1: EcoRI; Site 2:
 XhoI; The cDNA library was constructed from cells cultured
 in a medium with bubbling air containing 5% carbon
 dioxide"

ORIGIN

Alignment Scores:
 Pred. No.: 6.74e-55 Length: 478
 Score: 588.00 Matches: 116
 Percent Similarity: 86.7% Conservative: 1
 Best Local Similarity: 85.9% Mismatches: 0
 Query Match: 17.0% Indels: 18
 DB: 1 Gaps: 1

US-10-620-914-45 (1-648) x AV642589 (1-478)

Qy 1 MetGlySerGlyArgAspGlyArgProAlaSerTyrThrLysLysAsnPheSerLeuGlu 20
 Db 74 ATGGGTCTGGGTCTGTACGCCCGGCTCGAGGTACACCAAGAGAACTTCTCCCTGGAG 133
 Qy 21 LysLeuLysLeuSerSerMetLysAspAspLeuThrValLeuArgHisMetTrpPheGly 40
 Db 134 AAGCTCAAGCTCAGCAGCATGAAGGATGACCTGACCGCTTCTGCCCATATATGTGTGGC 193
 Qy 41 SerLysLysGlyAspAspHisAlaAlaArgLeuGluSerPheTyrGlyProGlnAlaAla 60
 Db 194 AGCAAGAAGGCGCATGATCAGCTGTCTGCTGCTGAGAGCTTCTACGGGCCCCCGCGCT 253

```

QY 61 AlaPhe----- 62
Db 254 GCCTATGATGCTTTCCGGTCGCGGTTCTCTGGGGTCGAGGCCCATGCTCGCTGCAGTT 313
QY 63 AlaAaArgLeuAlaGluArgSerAsnLeuIleTrrpValAspLeuGlyGlyThrGly 82
Db 314 GCTGCCCGCGCTGGCCGAGCGCTCGAACCTCATCTGGGTTGACCTGGGTGGTGCACCTGGG 373
QY 83 GluAsnValAspMetMetAlaAspTyrIleAspLeuAlaLysPheLysSerIleTyrVal 102
Db 374 GAGATGTCATATGATGCTGATTACATCGACTGGCGAAGTTCAAGTCCATCTACGTG 433
QY 103 ValAspLeuCysHisSerLeuCysGluValAlaLysLysAla 117
Db 434 GTCGACCTGTGCCACTCGCTGTGCGAGTGGCCAGGAAGGCG 478

RESULT 14
AV643824 480 bp mRNA linear EST 15-DEC-2000
LOCUS AV643824 Chlamydomonas reinhardtii 5% CO2 Chlamydomonas reinhardtii
DEFINITION CDNA clone HCL076g02_r 5', mRNA sequence.
ACCESSION AV643824 GI:10787152
VERSION AV643824.1
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii
ORGANISM Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;
Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.
REFERENCE 1 (bases 1 to 480)
AUTHORS Asamizu, E., Miura, K., Kucho, K., Inoue, Y., Fukuzawa, H., Ohyama, K.,
Nakamura, Y. and Tabata, S.
TITLE Generation of expressed sequence tags from low-CO2 and high-CO2
adapted cells of Chlamydomonas reinhardtii
JOURNAL DNA Res. 7 (5), 305-307 (2000)
PUBMED 11089912
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES
source
1..480
Location/Qualifiers
/organism="Chlamydomonas reinhardtii"
/mol_type="mRNA"
/strain="C9"
/db_xref="taxon:3055"
/clone="HCL076g02_r"
/notes="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2:
XhoI; The cDNA library was constructed from cells cultured
in a medium with bubbling air containing 5% carbon
dioxide"

ORIGIN
Alignment Scores:
Pred. No.: 1,91e-54 Length: 480
Score: 584.00 Matches: 115
Percent Similarity: 86.6% Conservative: 1
Best Local Similarity: 85.8% Mismatches: 0
Query Match: 16.9% Indels: 18
DB: 1 Gaps: 1

US-10-620-914-45 (1-648) x AV643824 (1-480)

QY 1 MetGlySerGlyArgAspGlyArgProAlaSerTyrThrLysLysAsnPheSerLeuGlu 20
Db 77 ATGGGGTGGGTGGTGGCGCGCGCTCGAGCTACACCAAGAAGAACTTCTCCCTGGAG 136
QY 21 LysLeuLysLeuSerSerMetLysAspAspLeuThrValLeuArgHisMetTrrpPheGly 40
Db 137 AGCTCAAGCTCAGCAGCATGAAGGATGACCTGACCGCTTCTGCGCCATATGTGTGGC 196

```

```

QY 41 SerLysGlyAspAspHisAlaAlaArgLeuGluSerPheTrrpGlyProGlnAlaAla 60
Db 197 AGCAAGAAGGGCGATGATCAGCTGCTCGCTGAGAGCTTCTACGGGCCCGAGCGCT 256
QY 61 AlaPhe----- 62
Db 257 GCCTATGATGCTTTCCGGTCGCGGTTCTCTGGGGTCGAGGCCCATGCTCGCTGCAGTT 316
QY 63 AlaAaArgLeuAlaGluArgSerAsnLeuIleTrrpValAspLeuGlyGlyThrGly 82
Db 317 GCTGCCCGCGCTGGCCGAGCGCTCGAACCTCATCTGGGTTGACCTGGGTGGTGCACCTGGG 376
QY 83 GluAsnValAspMetMetAlaAspTyrIleAspLeuAlaLysPheLysSerIleTyrVal 102
Db 377 GAGATGTCATATGATGCTGATTACATCGACTGGCGAAGTTCAAGTCCATCTACGTG 436
QY 103 ValAspLeuCysHisSerLeuCysGluValAlaLysLysAla 116
Db 437 GTCGACCTGTGCCACTCGCTGTGCGAGTGGCCAGGAAGG 478

RESULT 15
AV628989 466 bp mRNA linear EST 15-DEC-2000
LOCUS AV628989 Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas
DEFINITION reinhardtii cDNA clone LCL050a06_r 5', mRNA sequence.
ACCESSION AV628989
VERSION AV628989.1
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii
ORGANISM Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;
Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.
REFERENCE 1 (bases 1 to 466)
AUTHORS Asamizu, E., Miura, K., Kucho, K., Inoue, Y., Fukuzawa, H., Ohyama, K.,
Nakamura, Y. and Tabata, S.
TITLE Generation of expressed sequence tags from low-CO2 and high-CO2
adapted cells of Chlamydomonas reinhardtii
JOURNAL DNA Res. 7 (5), 305-307 (2000)
PUBMED 11089912
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES
Location/Qualifiers
1..466
/organism="Chlamydomonas reinhardtii"
/mol_type="mRNA"
/strain="C9"
/db_xref="taxon:3055"
/clone="LCL050a06_r"
/notes="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2:
XhoI; The cDNA library was constructed from cells cultured
in a carbon stress acclimatized condition in which carbon
dioxide concentration in the bubbling gas was changed from
5% to 0.04%"

ORIGIN
Alignment Scores:
Pred. No.: 2.5e-51 Length: 466
Score: 556.00 Matches: 109
Percent Similarity: 85.9% Conservative: 1
Best Local Similarity: 85.2% Mismatches: 0
Query Match: 16.1% Indels: 18
DB: 1 Gaps: 1

US-10-620-914-45 (1-648) x AV628989 (1-466)

QY 1 MetGlySerGlyArgAspGlyArgProAlaSerTyrThrLysLysAsnPheSerLeuGlu 20
Db 82 ATGGGGTGGGTGGTGGCGCGCGCTCGAGCTACACCAAGAAGAACTTCTCCCTGGAG 141

```

```
QY 21 LysLeuLysLeuSerSerMetLysAspAspLeuThrValLeuArgHisMetTrpPheGly 40
Db 142 AAGCTCAGCTCAGCAGCAAGGATGACCTGACCGTTCTGCGCCATATGTGTTGCGC 201
QY 41 SerLysLysGlyAspAspHisAlaAlaArgLeuGluSerPheTyrGlyProGlnAlaAla 60
Db 202 AGCAAGAAGGGCGATGATCAGCTGCTCGCTGGAGAGCTTCTACGGGCCCCAGGCGCT 261
QY 61 Alaphe----- 62
Db 262 GCCTATGATGCTTCCGGTCGCGGTTCTCTGCGGTCGCAGGCCCATGTGCTGCAGTT 321
QY 63 AlaAlaArgLeuAlaGluArgSerAsnLeuIleTrpValAspLeuGlyGlyThrGly 82
Db 322 GCTGCCCGCTGGCCGAGCGCTCGAACCTCATCTGGGTTGACCTGGGTGTGGCACTGGG 381
QY 83 GluAsnValAspMetMetAlaAspTyrIleAspLeuAlaLysPheLysSerIleTyrVal 102
Db 382 GAGAATGTCGATATGATGGCTGATTACATCGACCTGGCGAAGTTCAAGTCCATCTACGTG 441
QY 103 ValAspLeuCysHisSerLeuCys 110
Db 442 GTCGACCTGTGCCACTCGCTGTGC 465
```

Search completed: March 14, 2006, 06:44:05
Job time : 6860 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 14, 2006, 02:01:06 ; Search time 1035 Seconds
(without alignments)
4172.677 Million cell updates/sec

Title: US-10-620-914-45

Perfect score: 3463

Sequence: 1 MSGRDRPASVTKNFSLE.....RVNYSFFYMARKGAKDN 648

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Zgapop 6.0 , Zgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/abes/ABSSWEB spool/US10620914/runat_13032006_102038_6713/app_query.fasta_1
-DB=N Geneseq -QMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0 -LOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blom62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abes05h
-USER=US10620914 @CGN 1.1 727 @runat_13032006_102038_6713 -NCPU=6 -ICPU=3
-NO WMAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N_Geneseq_21:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002as:*
7: Geneseqn2002bs:*
8: Geneseqn2003as:*
9: Geneseqn2003bs:*
10: Geneseqn2003cs:*
11: Geneseqn2003ds:*
12: Geneseqn2004as:*
13: Geneseqn2004bs:*
14: Geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3463	100.0	1947	14	Adw72747 Chlamydom
2	2838	82.0	5242	14	Adw72746 Chlamydom
3	891	25.7	2259	12	Adp98536 C. albica
4	850.5	24.6	2727	14	Adw72752 Neurospor

5	828	23.9	2540	13	ADR85469	Adr85469 Aspergill
6	804	23.2	3427	14	ADM72751	Adm72751 Neurospor
7	700	20.2	2929	13	ADR84882	Adr84882 Aspergill
8	700	20.2	8929	13	ADR84295	Adr84295 Aspergill
9	379.5	11.0	1252	8	ABS58051	Abes58051 Gene enco
10	379.5	11.0	1252	14	ADM72704	Adm72704 Rhodobact
11	329	9.5	1248	8	ABS58076	Abss58076 Mesorhizo
12	329	9.5	1248	14	ADM72725	Adm72725 Mesorhizo
13	326.5	9.4	1251	8	ABS58082	Abss58082 Agrobacte
14	326.5	9.4	1251	14	ADM72731	Adm72731 Agrobacte
15	292.5	8.4	1251	8	ABS58084	Abss58084 Sinorhizo
16	292.5	8.4	1251	14	ADM72735	Adm72735 Sinorhizo
17	138.5	4.0	8832	14	ACL64521	ACL64521 M. xanthu
18	128	3.7	4246	6	ABS78845	Abbs78845 E. coli C
19	128	3.7	4246	10	ADH80412	Adh80412 Escherich
20	126.5	3.7	990	14	ACL72740	ACL72740 M. xanthu
21	123	3.6	7119	13	ADT04702	Adt04702 House mou
22	121	3.5	639	10	ADC93973	Adc93973 E. faeciu
23	120	3.5	666	8	ABS58085	Abss58085 Sinorhizo
24	120	3.5	666	14	ADM72737	Adm72737 Sinorhizo
25	119	3.4	6912	13	ADT04698	Adt04698 Human tum
26	119	3.4	6912	13	ADT04698	Adt04698 Human DNA
27	119	3.4	6912	14	ADZ49071	Adz49071 Insulin s
28	118.5	3.4	1155	3	AAA06598	Aaa06598 Human imm
29	118.5	3.4	1155	3	AAC81011	Aac81011 Human B11
30	118.5	3.4	1155	4	AHH93714	Ahh93714 Human pro
31	118.5	3.4	1155	4	AAI67211	Aai67211 B305D iso
32	118.5	3.4	1155	4	AA363807	Aas63807 Human pro
33	118.5	3.4	1155	4	AAH02779	Aah02779 Prostate
34	118.5	3.4	1155	4	AAH85028	Aah85028 Human pro
35	118.5	3.4	1155	5	ACA59615	Aca59615 Prostate
36	118.5	3.4	1155	6	ABL95178	Abi95178 Human B30
37	118.5	3.4	1155	6	AAS99857	Aas99857 Breast tu
38	118.5	3.4	1155	8	ACC95342	Acc95342 Prostate
39	118.5	3.4	1155	8	ADA11380	Adai1380 Human bre
40	118.5	3.4	1155	10	ADC15353	Adc15353 Human bre
41	118.5	3.4	1155	10	ADB13823	Adb13823 Human pro
42	118.5	3.4	1185	10	ADG26239	Adg26239 Human pro
43	118.5	3.4	1185	8	ADA11414	Adai1414 Human cdn
44	118.5	3.4	1185	10	ADC15387	Adc15387 Human bre
45	118.5	3.4	1383	8	ABZ79639	Abz79639 Mammaglob

ALIGNMENTS

RESULT 1

ADW72747
ID ADW72747 standard; cDNA; 1947 BP.

XX AC ADW72747;

XX DT 21-APR-2005 (first entry)

XX DE Chlamydomonas reinhardtii Btal cDNA.

XX KW Phospholipid synthesis; transgenic plant; fertilizer; ss; gene;
betaine lipid.

XX OS Chlamydomonas reinhardtii.

XX FH Key Location/Qualifiers
CDS 1..1947

FT FT /*tag= a
FT FT /product= "Btal protein"

XX PN WO2005009115-A2.

XX PD 03-FEB-2005.

XX PF 15-JUL-2004; 2004WO-US022789.

XX PR 16-JUL-2003; 2003US-00620914.

XX XX

PA	(UNMS) UNIV MICHIGAN STATE.	Db	421	ACCGGACGCTCATCACCCTTCTCTACTCGCTCACGATGATCCACCGTTCCACACGTC	480
XX	Benning C, Riekhof W;	Qy	161	IleAspGlnAlaCysSerTyrLeuSerGlnAspGlyLeuValGlyValAlaAlaAspPheTyr	180
XX	WPI; 2005-112975/12.	Db	481	ATCGACGAGCTTGCTGTAACCTGTCCTCCAGACGCGCTGGTGGCGCTTGCAGACTTCTAC	540
DR	P-PSDB; ADW72748.	Qy	181	ValSerGlyLysTyrAspLeuProLeuArgGlnMetProTyrSerArgPhePheTyr	200
XX	New composition comprising a purified DNA having an oligonucleotide	Db	541	GTGAGCGCAAGTACGACCTGCGCTGCGCCAGATGCCCTGGTGGCGCGTTCCTCTCGG	600
PT	sequence encoding a protein, useful in producing Betaine lipid compounds	Qy	201	ArgSerIlePheAspIleAspIleAspIleGlyProGluArgAlaTyrLeuGlu	220
PT	e.g., Diacylglycerol-1-O-4'-(N,N,N'-trimethyl) homoserine (DGTS).	Db	601	CGATCATCTTCGACATCGACATCGCCCGCGCGCGCGCGCTACCTCGGAG	660
XX	Claim 1; SEQ ID NO 44; 147pp; English.	Qy	221	GlnLysLeuGluArgValTyrGluGlnAsnThrGlnGlySerIleProTyrValProTyr	240
XX	The invention relates to a new composition comprising a purified DNA	Db	661	CAGAACTGAGCGCGCTGGGAGCAGAACACCCAGGGTTCGATCCCTACGTGGCGTG	720
CC	having an oligonucleotide sequence appearing as ADW72747 (Chlamydomonas	Qy	241	LeuArgAlaProTyrTyrValTyrIleGlyArgLeuProSerValGlyHisAlaLeuHis	260
CC	reinhardtii Btal cDNA) or ADW72752 (Neurospora crassa Btal coding region,	Db	721	CTGCGCGCCCTTACTACGTGTGGAATTGGCGCGCTGCCAGCGTTGGCCACGCCCTGCAC	780
CC	encoding Btal proteins which are DGTS (diacylglycerol-1-O-4'-(N,N,N'-	Qy	261	GluGluArgValGluArgProMetPheProProThrPheLeuTyrThrGlnSerTyr	280
CC	trimethyl)homoserine) synthetic enzymes. Btal (and BtaA and BtaB) enzymes	Db	781	GAGGAGCGGTGGAGCGCGCGCCCATGTTCGCCGCCACCTTCCTGTACACGCGAGTGTGG	840
CC	allow the replacement of phospholipids with non-phosphorus containing	Qy	281	GluAspProGluProAspMetGluValMetGluIleAsnProLysAspThrValLeuThr	300
CC	lipids in the cell membrane. A transgenic plant expressing the enzymes	Db	841	GAGGACCCGAGCGCGATATGGAGGTATGGAGATCAACCCCAAGACACGCGTGTGACC	900
CC	would have a reduced need for phosphate-containing fertilizer. Also	Qy	301	LeuThrSerGlyGlyCysAsnAlaLeuAsnLeuValGlnGlyAlaGlyGlnValVal	320
CC	included are an RNA transcribed from the purified DNA, antibodies	Db	901	CTGACTAGCGCGCGCTGCAATGCCCTGTGTCAGGGGCGCGCGCGAGTGGTG	960
CC	produced from the protein, a vector comprising the DNA or nucleic acid, a	Qy	321	SerValAspCysAsnProAlaGlnSerAlaLeuLeuGluLeuLysLysValAlaIleGln	340
CC	host cell comprising the vector, transgenic plants comprising the vector,	Db	961	TCGGTGGACTGCACACCCCGCGAGTCGCGCGCTTCCTGAGCTGAAGAGGTGGCGCATTCAG	1020
CC	and a protein translated from the RNA or encoded by the nucleic acid. The	Qy	341	GlnLeuGluPheGluAspValTyrGlnLeuPheGlyGlyValHisProArgIleGlu	360
CC	composition is useful in producing Betaine lipid compounds e.g.,	Db	1021	CAGCTGAGTGTGAGGACGTGTGCGAGCTGTTCGGGAGGCGGTGCACCCCGCATTTGAG	1080
CC	Diacylglycerol-1-O-4'-(N,N,N'-trimethyl) homoserine (DGTS). The composition	Qy	361	GluLeuTyrGluLysLysLeuAlaProPheLeuSerGlnThrSerHisAsnPheTyrSer	380
CC	of the invention reduces the amount of phosphate fertilizer needed for	Db	1081	GAGCTGTACGAGAAGAGTGGCGCGCTTCCTGCTGCTGCAACACGACCACTTCCTGTGTC	1140
CC	the optimal growth of crop plants. The present sequence encodes a betaine	Qy	381	LysArgLeuTyrTyrPheGlnHisGlyLeuTyrTyrGlnGlyMetGlyLysLeuCys	400
CC	lipid synthetic enzyme.	Db	1141	AAGCGCTCTGTGTACTTCCAGCAGCGCTGTACTACAGGGCGCATGGGCAAGCTGTGC	1200
XX	Sequence 1947 BP; 365 A; 619 C; 613 G; 350 T; 0 U; 0 Other;	Qy	401	TyrValLeuGlnCysLeuAlaValValLeuGlyLeuGlyLysThrValLysArgLeuAla	420
Alignment Scores:		Db	1201	TGGTGTCTGAGTGCCTTGGCGGTGTGTGGACTGGGCAAGACCTGCAAGCGCTCGCC	1260
Pred. No.:	0	Qy	421	AsnAlaProThrMetGluGluGlnArgArgLeuTyrAspSerAsnMetLeuIleHisPhe	440
Score:	3463.00	Db	1261	AACGCGCCCACTGGAGGAGCAGCGCGCTGTGTGGGACAGCAACATGCTCATCACTTC	1320
Percent Similarity:	100.0%	Qy	441	ValLysAsnGlyProLysProLeuValTyrPheValLysPheValSerLeuValLeu	460
Best Local Similarity:	100.0%	Db	1321	GTGAAGAACCGGCGCAAGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	1380
Query Match:	100.0%	Qy	461	PheAsnLysAlaValLeuTyrPheGlyGlyValProGlyLysGlnTyrAlaLeuIle	480
DB:	14	Db	1381	TTCAACAGCGCGT	1440
US-10-620-914-45 (1-648) x ADW72747 (1-1947)		Qy	481	LysAlaAspGlyIleProIleGluAsnTyrIleAlaArgThrMetAspGlyValAlaGlu	500
Qy	1 MetGlySerGlyArgAspGlyArgProAlaSerTyrThrLysLysAsnPheSerLeuGlu	Db	1441	AAGCGCAGCGCATCCCAATTGAGAACTACATCGCGCGCACCATGGAGCGCGTGGCGAG	1500
Db	1 ATGGGGTGGGTGTGACGGCGCGCGCTGCGAGCTACACCAAGAGAACTTCTCCTCGAG	Qy	501	AsnSerHisValArgLysGlnAsnTyrPheTyrTyrAsnCysLeuThrGlyLysPheLeu	520
Qy	21 LysLeuLysLeuSerSerMetLysAspAspLeuThrValLeuArgHisMetTyrPheGly	Db	1501	AACCTGCACTGCGCAAGCAGAACTACTTCTACTACAACTGCCTTCACCGCAAGTTCCTG	1560
Db	61 AAGCTCAAGCTCAGCAGCATGAAGATGACCTGACCGTTCGCGCCCATATGTGTTCCGC				
Qy	41 SerLysLysGlyAspAspHisAlaAlaArgLeuGluSerPheTyrGlyProGlnAlaAla				
Db	121 AGCAAGAAGGCGCATGATCAGCTGTCTCGCTGCGAGAGCTTCTACGGGCGCCCAAGCGCGCT				
Qy	61 AlaPheAlaAlaArgLeuAlaGluArgSerAsnLeuIleTyrValAspLeuGlyGly				
Db	181 GCCTTTGTGCGCGCGCTGGCGCGCGCTCGAACCTCATCTGGGTGTGAGTGTGGTGGC				
Qy	81 ThrGlyGluAsnValAspMetMetAlaAspTyrIleAspLeuAlaLysPheLysSerIle				
Db	241 ACTGGGAGATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG				
Qy	101 TyrValValAspLeuCysHisSerLeuCysGluValAlaLysLysLysAlaLysAlaLys				
Db	301 TACGTGTGCGACCTGTGCGACCTGTGCGAGTGGCGCGAGAGAGAGAGAGAGAGAGAG				
Qy	121 GlyTyrLysAsnValGlnValValGluAlaAspAlaCysGlnPheAlaProProGluGly				
Db	361 GGCTGGAGAGATGTCCAGGTGTGGAGCGCGACCTTGCCTAATTTGGCCCTTGAAGGC				
Qy	141 ThrAlaThrLeuIleThrPheSerTyrSerLeuThrMetIleProProPheHisAsnVal				

QY 521 ArgAspAsnCysProThrThrValSerThrValArgGluAlaAlaPheAlaThrLeuLysSerGlyVal 540
 DB 1561 CGCGACAACCTGCCCACTACCTGCGCGAGCGCGCTTCGCCACCTCAAGAGTGGCGTG 1620
 QY 541 ValAspAsnLeuThrValSerThrAsnPhePheMetClnGluLeuLysAlaArgThrTyr 560
 DB 1621 GTGGACAACCTGACCGCTCCACCACTTCTTATGGAGGAGCTCAAGGCGCACCTAC 1680
 QY 561 ThrLysValLeuMetAspHisValAspTyrLeuAspMetProValAlaAsnGluLeu 580
 DB 1681 ACCAAGTGATTCGTGGACCACTGAGCTGGCTGGATATGCCGTGGCCACAGAGCTG 1740
 QY 581 AlaGluCysLeuAlaLysGlnValAlaProGlyGlyLeValIleTyrArgSerAlaSer 600
 DB 1741 GCCGAGTGCTGGCCCAAGCAGGTGGCGCGCGCGGCGATCATCTGGCGCTCCGCGCTCC 1800
 QY 601 LeuSerProProThrAlaGluLeuIleGlnLysAlaGlyPheAspValAlaArgCysIleArg 620
 DB 1801 CTGAGCGCGCGCTTACCGCGAGCTGATCCAGAGCGCGGCTTCGACGTGGCGCTGCATCCGC 1860
 QY 621 ArgAlaThrGlnGlyTyrMetAspArgValAsnMetTyrSerSerPheTyrMetAlaArg 640
 DB 1861 CGCGCCACTCAGGCTACATGGACCGCTCAACATGTACAGCTCCTTACATGGCGCGC 1920
 QY 641 ArgLysGlyAlaLysLysAspAsn 648
 DB 1921 CGGAAGGCGGCCAAGAAGGACAAC 1944

RESULT 2

ADW72746
 ID ADW72746 standard; DNA; 5242 BP.

XX AC
 AC ADW72746;

DT 21-APR-2005 (first entry)

XX Chlamydomonas reinhardtii Btal gene.

XX Phospholipid synthesis; transgenic plant; fertilizer; ds; gene;
 KW betaine lipid.

XX Chlamydomonas reinhardtii.

OS WO2005009115-A2.

PN 03-FEB-2005.

XX 15-JUL-2004; 2004WO-US022789.

PR 16-JUL-2003; 2003US-00620914.

XX (UNMS) UNIV MICHIGAN STATE.

XX Benning C, Riekhof W;

PI WPI; 2005-112975/12.

DR P-PSDB; ADW72748.

XX New composition comprising a purified DNA having an oligonucleotide
 PT sequence encoding a protein, useful in producing Betaine lipid compounds
 PT e.g., Diacylglycerol-1-O'-4'-(N,N,N-trimethyl) homoserine (DGTS).

XX Example 4; SEQ ID NO 43; 147pp; English.

XX The invention relates to a new composition comprising a purified DNA
 CC having an oligonucleotide sequence appearing as ADW72747 (Chlamydomonas
 CC reinhardtii Btal cDNA) or ADW72752 (Neurospora crassa Btal coding region,
 CC encoding Btal proteins which are DGTS (diacylglycerol-1-O'-4'-(N,N,N-
 CC trimethyl)homoserine) synthetic enzymes. Btal (and BtaA and BtaB) enzymes
 CC allow the replacement of phospholipids with non-phosphorus containing
 CC lipids in the cell membrane. A transgenic plant expressing the enzymes
 CC would have a reduced need for phosphate-containing fertilizer. Also

CC included are an RNA transcribed from the purified DNA, antibodies
 CC produced from the protein, a vector comprising the DNA or nucleic acid, a
 CC host cell comprising the vector, transgenic plants comprising the vector,
 CC and a protein translated from the RNA or encoded by the nucleic acid. The
 CC composition is useful in producing Betaine lipid compounds e.g.,
 CC Diacylglycerol-1-O'-4'-(N,N,N-trimethyl) homoserine (DGTS). The composition
 CC of the invention reduces the amount of phosphate fertilizer needed for
 CC the optimal growth of crop plants. The present sequence encodes a betaine
 CC lipid synthetic enzyme.

XX SQ Sequence 5242 BP; 955 A; 1562 C; 1662 G; 1063 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1.83e-296 Length: 5242
 Score: 2838.00 Matches: 645
 Percent Similarity: 45.6% Conservatives: 0
 Best Local Similarity: 45.6% Mismatches: 3
 Query Match: 82.0% Indels: 771
 DB: 14 Gaps: 11

US-10-620-914-45 (1-648) x ADW72746 (1-5242)

QY 1 MetGlySerGlyArgAspGlyArgProAlaSerTyrThrLysLysAsnPheSerLeuGlu 20
 DB 501 ATGGGGTTCGGGTTCGTGACGCGCGCTTCGAGGTACACCAAGAAGAACTTCTCCCTGGAG 560
 QY 21 LysLeuLysLeuSerSerMetLysAspAspLeuThrValLeuArgHisMetTyrPheGly 40
 DB 561 AAGCTCAAGCTCAGCAGCATGAAGATGACCTGACCGCTTCGCCCATATGTGTTCCGC 620
 QY 41 SerLysGlyAspAspHisAlaAlaArgLeuGluSerPheTyrGlyProGlnAlaAla 60
 DB 621 AGCAAGAGGCGGATGATCAGCTGCTGCGCTGCGAGAGCTTCTACGGGCGCCAGCGCGCT 680
 QY 61 AlaPhe----- 62
 DB 681 GCCTG-TAAGTCGCGATTAACTTTATGCTCATTTAAGTGTTCGAGTTGCTTACGCGCTG 739
 QY 62 ----- 62
 DB 740 TCTGCCCGCGCAGATGATGCTTTCCGGTTCGGGTTCCTCTGCGGTTCGAGCGCCATGCTC 799
 QY 63 -----AlaAlaArgLeuAlaGluArgSerAsnLeuIleTyrValAspLeuGlyGly 79
 DB 800 GTGCAGTTGCTGCGCGCTGCGCGAGCGCTCGAACCTCATCTGCGGTTCACCTGGGTGGT 859
 QY 80 GlyThrGly----- 82
 DB 860 GGCACCTGGGGTGAGTCGCGGTGTATCGGGGAGGTGCGGATAGCCCTGCGCGCTTGGCT 919
 QY 83 -----GluAsnVal 85
 DB 920 GCGCCTCGCGGCCATTATTATGCTGATCTGCGCTTTCCTTTGTCGCGCAGGAGATGTC 979
 QY 86 AspMetMetAlaAspTyrIleAspLeuAlaLysPheLysSerIleTyrValValAspLeu 105
 DB 980 GATATGATGCTGATTACATCGACCTGCGGAGTTCAGTCCATCTACGTGGTGCAGCTG 1039
 QY 106 CysHisSerLeuCysGluValAlaLysLysAlaLysAlaLysGlyTyrLysAsnVal 125
 DB 1040 TGCACCTGCTGTCGAGGTGGCCAAAGAGAGCGCAAGGCGCAAGGCTCGAAGATGTC 1099
 QY 126 GlnValValGluAlaAspAlaCysGlnPheAlaProGluGlyThrAlaThrLeuIle 145
 DB 1100 CAGGTCTGGAGCGCGCTGTCCTTTCGCAATTTGCGCCCTTCGAGGCGCAGCGCTCATC 1159
 QY 146 ThrPheSerTyrSerLeuThrMet----- 153
 DB 1160 ACCTTCTCTACTCTGCTCAGAG-TGAGTTGCAACCGCTCGACTTGCATCGGAGGATC 1218
 QY 154 -----IlePro 155
 DB 1219 CATCCACCACCAACAGTTTCATCCCTCTCACCCCGCGCTTTTGTGTTGCGAGTGAATCCA 1278

156 ProPheHisAsnValIleAspGlnAlaCysSerTyrIleuSerGlnAspGlyLeuValGly 175
1279 CCGTTCACAAAGCTCATCGACAGGCTTGTCTGTACCTGTCCCAAGACGGCCCTGGTGGC 1338
176 ValAlaAspPheTyrValSerGlyLysTyrAspLeuProLeuArgGlnMetProTrpSer 195
1339 GTTGGCGACTTCTACGTGAGCGGCAAGTACGACCTGCCCCCTGGCCAGATGCCCTGGTCG 1398
196 ArgArgPhePheTrp 200
1399 CGCCGTTTCTTCTGGCGGTGAGTTACCAAGCGCGGCTACACATCTCTCAAGCAAGCGTA 1458
200 200 200
1459 GCTGAGCGGGCACAGAAAGCGGTAGGGCCCGGGTGCCTTCTGGGTAGTGGCAAGGA 1518
200 200 200
1519 TATCGGCAAGGCCCTTCAGCTACCCGAGGTACCATGGTGATCTGTCAAGCAGCTCATCAC 1578
200 200 200
1579 GGTGCTGACTGAGGCACGTGCTCGAGTCTCGGTCTGTATAGGGCTTTCAGCACATCACA 1638
200 200 200
1639 GCGAGTCTCCGAATGCGGTGCTGACAGGTGTATGCTCGGGACAGCACCATGCACCGT 1698
200 200 200
1699 GCGGCGTTCTTGACTTCATCGCTGCTCTCGGGCCCTCCCTGGGTTTCGCTCAGACACGT 1758
201 201 201
1759 CTCGCTGCCCTGTGCTGTGCTGCTGTGTCAGATCGATCTTCGACATCGCAAC 1818
209 IleAspIleGlyProGluArgAlaTyrIleuGlnLysLeuGluArgValTrpGlu 228
1819 ATTGACATCGCCCCGAGCGCGGCTTACTGAGCAGAAGCTGGAGCGGTGGAG 1878
229 GlnAsnThr 231
1879 CAGAACAACCCAGGTGTGTGCGGCTTGACACCTGGCCTGTGCAAGCCCAAGTAAAGTTTG 1938
231 231 231
1939 CTCATCAGCCCTGCAGCACCCGCGGTGCGGGGATATTATCGGCAGACCGTCCGCA 1998
232 232 232
1999 CTGTTGGTTAATCCACAGTCCGCTGCTTCTCTGCGTGTTCAGGGTTCGATCCCC 2058
237 TyrValProTrpLeuArgAlaProTyrTyrValTrpIleGlyArgLeuProSerValGly 256
2059 TACGTGCGGTGGCTGCGCGCCCCCTACTAGTGTGGATTGCGCGCTGCGCAGCGTTGGC 2118
257 His 257
2119 CG-TGAGTACGCTGCGCGGGACCTCTGCGCATACAGGACACATGTTACGGGGGCTCA 2177
257 257 257
2178 GTGAAGCGTATCGTATCGTTGACGTCTGTCTCAAGGTACCAAGCGCGTCCCTTG 2237
257 257 257
2238 CGCGTTGAAATTGTTGGCGCAAGCATGAGCGGTGCGCGCATGCCCACTCGCGTTGAC 2297
258 258 258
2298 CCTGCAGACGCCCCGACGAGGCGGTGAGCGCGGCCCATGTTTCCCGCCCACTTC 2357

275 LeuTyrThrGlnSerTrpGluAspProGluProAspMetGlu 288
2358 CTGTACACGCACTCGTGGGAGGACCCCGAGCCGATATGGAGCTGAGCGGCTCAGCCTGA 2417
288 288 288
2418 GGCAGATACGCACAAAGAAACGCTCGTGTGTACCCGAGGGCGGATACACGACGTTTCG 2477
288 288 288
2478 AAGCATTTGGGCACCGTCACTCGTCGACCAAGCAAGCAGCACTAAGCACCAGCGGTT 2537
288 288 288
2538 CTGTGGCAATGTGCCACCGCAATGCTGSCCAGTGCAGCTTCGCAATTAGTGTATGCC 2597
288 288 288
2598 AGTATCACGCTTAGCTCAGCTGCAAGCTCTGTGTAGAAAGCAGCGATGTGGCACT 2657
289 289 289
2658 GACGGCAACACTTCCTGCTGAAACTGTGCTGCCATCCTCTCGTGCCCTTCCAGGTGAT 2717
290 tGluIleAsnProLysAspThrValLeuThrLeuThrSerGlyGlyCysAsnAlaLeuAs 310
2718 GGAGATCAACCCCAAGSACACGCGTGTGACCTGACTAGCGCGGCTGCAATGCCCTGAA 2777
310 nLeuLeuValGlnGlyAlaGly 317
2778 CCTGCTGCTGACGGGGCGCGCAGGTGAGATCATAGTGGTCTGTCACCTTCTTGGTG 2837
317 317 317
2838 TCCTTACTTCATGATGCCCGCTGCACGGAATGTTAGAAAGCGGTGGCATACATGTAGCT 2897
318 318 318
2898 GCAGAATGCTGCTGACGTCTCTTCGCGCTGCTGTGTCAGGTGGTGTGCTGGTACTGCA 2957
325 snProAlaGlnSerAlaLeuLeuGluLeuLysValAlaIleGlnGlnLeuGluPheG 345
2958 ACCCGCGCAGTGGCGCTTCTGAGCTGAAGAAGGTGGCCATTCACGACGTGGAGTTG 3017
345 LuAspValTrp 348
3018 AGGACGTGTGGCAGGTAAGGGGCTCTCATCTGGCGCTAGGACTGCACACGCTCTGTC 3077
348 348 348
3078 CTACTGTGATGCAACATGCGGACCTGGTCAGGCTTTCGCGCATTCGCACTGCGATGCTG 3137
348 348 348
3138 TCCCCCAGGGCTCAGCACATCTCTCGCGGCTGCAACGGCAGACGCCCTTTTACACTCC 3197
349 349 349
3198 GCAACCCCTCTCTTGGTAATTCCCTTCCACAGCTGTTCGCGAGGGCGTGCACCCGCGC 3257
359 IleGluGluLeuTyrGluLysLysLeuAlaProPheLeuSerGlnThrSerHisAsnPhe 378
3258 ATTGAGGAGCTGTACGAGAAGAAGTGGCGCCCTTCTGTCGCAAAACAGCAGCAACTTC 3317
379 TrpSerLysArgLeuTrpTyrPheGlnHisGlyLeuTyrTyrGlnGlyGlyMetGlyLys 398
3318 TGTTCGAAGCGCTCTGTACTTCCAGCAGCGCTGTACTACAGGGCGGATGGGCAAG 3377
399 LeuCysTrpValLeuGlnCysLeuAlaValValLeuGlyLeuGlyLysThrValLysArg 418
3378 CTGTGCTGGTGTGTCAGTGCCTGGCGTGTGCTGGAGTGGGCAAGACCGTCAAGCGC 3437
419 LeuAlaAsnAlaProThrMetGluGluGlnArgLeuTrpAspSerAsnMetLeuIle 438

Db 3438 CTGCCAACCGCCCAACATGGAGGACGCGCGCTGTGGGACAGCAACATGCTCATC 3497
Qy |||||||
Db 439 HiePheVallyAsnGlyProLysProLysProLysPheVallyPheValSerLeu 458
Qy |||||||
Db 3498 CACTTCGTGAAGAACGGCCCGCCAGCCCTGGTGGCTGTTCGTCAAGTTCTGTAGCGTG 3557
Qy |||||||
Db 459 ValLeuPheAsnLysAlaValLeuTyrPheGlyGlyValProGlyLysGlnTyrAla 478
Qy |||||||
Db 3558 GTGCTCTTCAACAGGCGGTGTGTGTTCGGCGCGCGGTGCGCGGCAAGCAGTACGCG 3617
Qy |||||||
Db 479 LeuIleLysAlaAspGlyIleProIleGluAsnTyrIleAlaArgThrMetAspGlyVal 498
Qy |||||||
Db 3618 CTGATCAAGCGGACGATCCCATTTGAGAACTTACATCGCGCGCACCATGGACGCGGTG 3677
Qy |||||||
Db 499 AlaGluAsnSerHisValArgLysGlnAsnTyrPheTyrTyrAsnCysLeuThrGlyLys 518
Qy |||||||
Db 3678 GCGGAGAACTCGACGTCGCGACGAGCAACTACTTCTACTACAACTGCTCAGCGGCAAG 3737
Qy |||||||
Db 519 PheLeuArgAspAsnCysProThrTyrLeuArgGluAlaPheAlaThrLeuLysSer 538
Qy |||||||
Db 3738 TTCTTGGCGACAACTGCCCACTACCTGCGGAGCGGCTTGGCACCTCAAGAGT 3797
Qy |||||||
Db 539 GlyValValAspAsnLeuThrValSerThrAsnPheMetGluGluLeuLysAlaArg 558
Qy |||||||
Db 3798 GCGGTGTGTGACAACTGACCGCTTCCACCACTTCTTCATGAGGAGGCTCAAGAGCGCG 3857
Qy |||||||
Db 559 ThrTyrThrLys 562
Qy |||||||
Db 3858 ACCTACCAAGGTGGGAGCCAGCGGAGAGTGTGAAAGCGGAGATGCGCGGCGTTG 3917
Qy |||||||
Db 562 562
Db 3918 TAGGTGCATGCGGAACGCTTGTCTGCCAGGTTGAAAGGCGAGCGAGGAAACGGAATT 3977
Qy |||||||
Db 562 562
Db 3978 TCAGAAATTCGGATGCTGGGAAGCGGAACGAAACAAAGAACGGAATAGAACGCGCTGGCG 4037
Qy |||||||
Db 562 562
Db 4038 AGAACTGGTGGTGGCTTTGGCATCCAGTCTGCCAGGCGGCGCGCTACCTGCT 4097
Qy |||||||
Db 563 Val 563
Qy |||||||
Db 4098 CAGTTCGCTTAACCGTTACAACTCGCCCTCACCGCCCCCGCCGCCACCGAGTG 4157
Qy |||||||
Db 564 IleuMetAspHisValAspTyrLeuAspMetProValAlaAsnGluLeuAlaGluCys 583
Qy |||||||
Db 4158 ATTCTGATGACCAACGCTGGACTGGCTGGATATGCCCGTGCCCAACGAGCTGGCGGATGC 4217
Qy |||||||
Db 584 LeuAlaLysGlnValAlaProGlyGlyIleValIleTyrArgSerAlaSerLeuSerPro 603
Qy |||||||
Db 4218 CTGGCCNAGCAGGTGGCGCGGGCGGATCGTATCTGGCGCTCCGCTCCCTCAGCCCG 4277
Qy |||||||
Db 604 ProTyrAlaGluLeuIleGlnLysAlaGlyPheAspValArgCysIleArgAlaThr 623
Qy |||||||
Db 4278 CCCTAGCCCGAGCTGATCCAGAGGCGGCTTCGAGTGGCTGCATCCGCGCGGCACT 4337
Qy |||||||
Db 624 GlnGlyTyrMetAspArgValAsnMetTyr 633
Qy |||||||
Db 4338 CAGGGCTTACATGACCGCGTCAACATGTATACAGTGGCGTGTCTTTGTCGAGGTGTGAG 4397
Qy |||||||
Db 633 633
Db 4398 GATGTAGGATGTCAAGTGGTGGCGGAGGAGCATGCGGTCGAATGGGAGGAGTTGCGT 4457
Qy |||||||
Db 633 633
Db 4458 TATTGATCCCGATGGACCTGGTGAATGCGGTGAGCGGCTACCGTATGTGCCAAGGAC 4517
Qy |||||||
Db 633 633

Db 4518 GGGGTTGGCGCGCTCGGCGCGCAGCAGAAACTTTGAAAGGTTGTTTATAGATCAAGATA 4577
Qy 633 633
Db 4578 TGTTGCGATTAACGGAAGCGACCTAGCGTCAGGGCGCAACAGCGGTGGCACCAGGAA 4637
Qy 634 634
Db 4638 TCTGCGTTGTCTTCGATGTATACCTGTGATCCCAATCTTGTGCGTTTCTCCAATGCGCAGC 4697
Qy Ser 634
635 SerPheTyrMetAlaArgArgLysGlyValAlaLysIleAspAsn 648
6498 TCCTTCTACATGCGCGCGCGGAAAGGCGCCCAAGAGGACAAAC 4739
RESULT 3
ADP98536
ID ADP98536 standard; DNA; 2259 BP.
XX ADP98536;
AC ADP98536;
XX 23-SBP-2004 (first entry)
DT XX
XX C. albicans specific gene, orf6.3438, DNA sequence.
DE XX
XX Diploid fungal cell; allele; gene disruption cassette;
KW promoter replacement fragment; antifungal; fungicide; gene therapy;
KW infection; Candida albicans; gene; ds.
XX
OS Candida albicans.
XX
FN WO2004056965-A2.
XX
PD 08-JUL-2004.
XX
PF 19-DEC-2003; 2003WO-US040618.
XX
XX 19-DEC-2002; 2002US-0434832P.
PR
XX (ELIT-) ELITRA PHARM INC.
PA (ELIT-) ELITRA CANADA LTD.
XX
XX Roemer T, Jiang B, Boone C, Bussey H;
PI
XX WPI; 2004-500296/47.
DR P-PSDB; ADP98846.
XX
XX Constructing a strain of diploid fungal cells in which both alleles of a
PT gene are modified comprises modifying the alleles of a gene in the fungal
PT cells by recombination using a gene disruption cassette and a promoter
PT replacement fragment.
XX
Claim 33; SEQ ID NO 6021; 163pp; English.
PS
XX The invention relates to a novel method for constructing a strain of
CC diploid fungal cells in which both alleles of a gene are modified. The
CC method comprises modifying the alleles of a gene in diploid fungal cells
CC by recombination using a gene disruption cassette and a promoter.
CC replacement fragment. The invention further comprises: assembling a
CC collection of diploid fungal cells each of which comprises modified
CC alleles of a different gene; a strain of diploid fungal cells comprising
CC modified alleles of a gene, where the first allele of the gene is
CC inactivated by a gene disruption cassette comprising a nucleotide
CC sequence encoding an expressible selectable marker; and the expression of
CC the second allele of the gene is regulated by a heterologous promoter
CC that is operably linked to the coding region of the second allele of the
CC gene, and where the gene encodes the polypeptide mentioned above; a
CC collection of diploid fungal strains comprising the diploid strains cited
CC above, where substantially all the different genes that encode the above
CC amino acid sequences are modified and are present in different diploid
CC strains in the collection; a nucleic acid molecule comprising a
CC nucleic acid molecules, where each nucleic acid molecule comprises a
CC nucleotide sequence that is hybridizable to a target nucleotide sequence
CC comprising any of the 310 nucleotide sequences listed in the

CC specification (ADP98516-ADP98825); identifying a gene that is essential
 CC to the survival or growth of a fungus, that contributes to the virulence
 CC and/or pathogenicity of a fungus, or that contributes to the resistance
 CC of a diploid fungus to an antifungal agent; identifying an antifungal
 CC agent that inhibits the growth of a diploid fungus, or a therapeutic
 CC agent for treatment of a mammalian disease; correlating changes in the
 CC levels of proteins or gene transcripts with the inhibition of growth or
 CC proliferation of a diploid fungal cell; a purified or isolated nucleic
 CC acid molecule comprising a nucleotide sequence encoding a gene product
 CC required for proliferation of *Candida albicans*, where the gene product
 CC consists of any of the above-mentioned amino acid sequences; a vector
 CC comprising a promoter operably linked to the nucleic acid molecule cited
 CC above; a host cell containing the vector; a purified or isolated
 CC polypeptide comprising any of the 61 amino acid sequences given in the
 CC specification (ADP96718-ADP96778); a fusion protein comprising a fragment
 CC consisting of at least 6 consecutive residues of any of ADP98826-ADP99135
 CC ; producing a polypeptide; identifying a compound which modulates the
 CC activity of a gene product encoded by a nucleic acid comprising any of
 CC ADP98516-ADP98825; eliciting an immune response in an animal; a strain of
 CC *Candida albicans*, where a first allele of a gene comprising any of
 CC ADP98516-ADP98825 is inactive and a second allele of the gene is under
 CC the control of a heterologous promoter; identifying a compound or binding
 CC partner that binds to the polypeptide comprising any of ADP98826-
 CC ADP99135, or its fragment; identifying a compound having the ability to
 CC inhibit growth or proliferation of *Candida albicans*; inhibiting growth or
 CC proliferation of *Candida albicans* cells; manufacturing an antimycotic
 CC compound; treating an infection of a subject by *Candida albicans*;
 CC preventing or containing contamination of an object by *Candida albicans*,
 CC or for preventing or inhibiting formation on a surface of a biofilm
 CC comprising *Candida albicans*; a pharmaceutical composition comprising a
 CC therapeutic amount of an agent which reduces the activity or level of a
 CC gene product encoded by a nucleic acid comprising any of ADP98516-
 CC ADP98825 in a pharmaceutical carrier; an antibody preparation which binds
 CC to the polypeptide; methods for evaluating a compound against a target gene
 CC product encoded by any of ADP98516-ADP98825; identifying an antimycotic
 CC compound; a computer or a computer readable medium that comprises at
 CC least one of the nucleotide sequences mentioned in the specification or
 CC at least one amino acid sequence selected from ADP98826-ADP99135; a
 CC method assisted by a computer for identifying a putatively essential gene
 CC of a fungus; and a protein array comprising proteins, where at least one
 CC protein comprises an amino acid sequence or a portion of an amino acid
 CC sequence selected from ADP98516-ADP98825. The novel methods and
 CC compositions have fungicide activity. The compositions may be used in
 CC gene therapy. The composition and methods are useful for drug screening
 CC purposes or for diagnosing, preventing or treating infections associated
 CC with *Candida albicans*. These may also be used for constructing strains
 CC useful for identification and validation of gene products as effective
 CC targets for therapeutic intervention, for identifying and validating gene
 CC products as effective targets for therapeutic intervention, and for
 CC collecting identified essential genes. This polynucleotide sequence
 CC represents a *Candida albicans* fungal specific gene of the invention.
 CC NOTE: This sequence was downloaded from an electronic sequence listing
 CC provided on the WIPO website.

XX Sequence 2259 BP; 714 A; 359 C; 406 G; 780 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1.12e-85	Length:	2259
Score:	891.00	Matches:	229
Percent Similarity:	46.6%	Conservative:	98
Best Local Similarity:	32.7%	Mismatches:	232
Query Match:	25.7%	Indels:	142
DB:	12	Gaps:	21

US-10-620-914-45 (1-648) x ADP98536 (1-2259)

Qy	51	LeuGluSerPheTyrGlyProGlnAlaAlaAlaPhe-----	62
Db	208	TTAGAGTCATCTACAGAAATCAAGCTCATATTTATGATAACACATAGAGAAATTTTGTG	267
Qy	63	-----AlaAlaArgLeuAlaGluArgSerAsnLeu 72	

Db	268	AAAGTAGACAGAAATGCTTAGATTAGCTATTATCCCAATTTCCCAAAAAAAGATCTT	327
Qy	73	ILeTrrValAspLeuGlyGlyThrGlyGluAsnValAspMetMetAlaAspTyrIle	92
Db	328	ATTGGATTGATATGGTGGAACTGGTTCCAATATTGAATTCATGGATGAATAGT	387
Qy	93	AspLeuAlaLys---PheLysSerIleTyrValValAspLeuCysHisSerLeuCysGlu	111
Db	388	AAATATCTGAAACCTTTAAAGCTGTTATTGTTGGTATCTTTCCCATCTTTGTGTGAA	447
Qy	112	ValAlaLysLysLysAlaLysGlyTrrPlyAsnValGlnValValGluAlaAsp	131
Db	448	GTTGCTAAGGCAAGATTGAAGCCCATGAATGTCATGATTTATGTTGCTGAT	507
Qy	132	AlaCysGlnPheAlaProGluGlyThrAlaThrLeuIleThrPheSerTyrSerLeu	151
Db	508	GCCTGTGATTTTACTATTGATTATGATGCTGCTGATTTGATTTCTTCTTATTCATTG	567
Qy	152	ThrMetIleProProPheHisAsnValIleAspGlnAlaCysSerTyrLeuSerGlnAsp	171
Db	568	TCGATGATCCCACTTTCATGCTGCTATCGATAATGCTGTTCTTAATTAGATATGAA	627
Qy	172	GlyLeuValGlyValAlaAspPheTyrValSer-----GlyLysTyr	185
Db	628	GGTATTATTATCCCACTGGGATTTTGGTATTCAAAGCAGTCACACCTCAATGGGTCGATC	687
Qy	186	Asp-----LeuProLeuArgGlnMetProTrrPrrSerArgArgPhePrrArg	201
Db	688	AATACTGTTGGTGGTGGTTACAGGGACATTCCTTGGATATACGTATATTTTGGAGA	747
Qy	202	SerIlePheAspIleAspAsnIleAspIleGlyProGluArgArgAlaTrrLeuGluGln	221
Db	748	ATTGGTTTGAAGCTGATAAAGTGTGTTGGATCTTCTCAAGAGAAGAACTATTGGAATAT	807
Qy	222	LysLeuGluArgValTrrPrrGluGlnAsnThrGln-----GlySerIleProTrr	237
Db	808	AAATTTGGTACCGCTCAAACTTTTGAATCTTACAACAGGCTTTGGGTAAGT-----	861
Qy	238	ValProTrrLeuArgAlaProTrrTrrValTrrPrrIleGlyArgLeuProSerValGlyHis	257
Db	862	-----CCCTATTATTATTTGGATGGTGTGATTAATCAAAATCAACAC	903
Qy	258	AlaLeuHisGluGlu-----ArgValGluArgProProMetPheProProTrr	273
Db	904	ACCATTTTAGAAGATGGAATTTGTTAGCCACTGAATCCCTTACCTTGTCTCAACATCA	963
Qy	273	-----	273
Db	964	ACTCCAATCGCTAATCAACTTGAAGATATTCCAATTTCTAAAGTCATGAAGTCGCTTTA	1023
Qy	273	-----	273
Db	1024	ATCACTTGCACAAAAAATTTACCTTACCCTCAATCAATGACTATCAAAAGGAATATTGGA	1083
Qy	274	-----PheLeuTrrThr	277
Db	1084	GTCTACTATGATGAATGAATCCATGTATGAACAATTTAAAAACCAATACATTTATGCT	1143
Qy	278	GlnSerTrrPrrGluAspProGluProAspMetGluValMetGluIleAsnProLysAspThr	297
Db	1144	TTCACTTGGGAAGATCTCTCGTGAAGATCATATAAUAUAUAUAUAUAUAUAUAUAUAUA	1203
Qy	298	ValLeuThrLeuThrSerGlyGlyCysAsnAlaLeuAsnLeu-----LeuValGlnGly	315
Db	1204	GTTTGGCTATTACTTCAGCTGGTGATATAATTTTGAGTATGCTAGTTTACCAACACCA	1263
Qy	316	AlaGlyGlnValValSerValAspCysAsnProAlaGlnSerAlaLeuLeuGluLeuLys	335
Db	1264	CCAAAAAGATTCATGCTGTTGATCTTAATCCATGTCATAAACCACATTTATTAGAAATG	1323
Qy	336	LysValAlaIleGlnGlnLeuGluPheGluAspValTrrPrrGlnLeuPheGlyGluVal	355
Db	1324	TTGGCTAGTTTATGATGCTCTTCTCAAGAAACAATTTGGTCAATGTTTGGTGAAGGTAA	1383

Qy 356 HisProArgIleGluGluLeuTyrGluLysLeuAlaProPheLeuSerGlnThrSer 375
 Db 1384 ATCGAAATTTCAATGATCTTTTGGTATGATCTTTGGCACCACATGCTCTCTAATGCC 1443
 Qy 376 HisasnPheTrpSerLysArg-----LeuTrpTyrPheGlnHisGlyLeuTyrGln 393
 Db 1444 TTCCAACTAGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1497
 Qy 394 GlyGlyMetGlyLysLeuCysTrpValLeuGlnCys-----LeuAlaValValLeu 410
 Db 1498 ---GATACCTGGTCTTCTAGATGGCATTAAAGTATATCAAGATATGTTTCAAGTTGC 1554
 Qy 411 GlyLeuGlyLysThrValLysArgLeuAlaAsnAlaProThrMetGluGlnArgArg 430
 Db 1555 GGTGTTAGTAAATACGTTGAAGAACTTTGTCTGCAACCACTGGAAGCAACATTGAGA 1614
 Qy 431 LeuTrpAspSerAsnMetLeuIleHisPheValLysAsnGlyProLysProLeuValTrp 450
 Db 1615 ATTTGGAAATGAACATTTG-----AAACCAACT----- 1641
 Qy 451 LeuPheValLysPheVal----SerLeuValLeuPheAsnLysAlaValLeuTrpPheGly 469
 Db 1642 TTATTTAATCCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1701
 Qy 470 GlyGlyValProGlyLysGlnTyrAlaLeuLysAlaAspGlyIleProIleGluAsn 489
 Db 1702 TTGGGAGTTCCAGTAAATCAAGCAGCTTTAATG-----GGACCATCTGTTATTAAA 1752
 Qy 490 TyrIleAlaArgThrMetAspGlyValAlaGluAsnSerHisValArgLysGlnAsnTyr 509
 Db 1753 TATGTTGTTGATCTTTGGACCAATCAATTAAGAGATCGATGATTTCAATATGATACTAT 1812
 Qy 510 PheTyrTyrAsnCysLeuThrGlyLysPheLeuArgAspAsnCysProThrTyrLeuArg 529
 Db 1813 TTCTACTATTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1872
 Qy 530 GluAlaAlaPheAlaThrLeuLys-----SerGlyValVal 541
 Db 1873 ACAAAAGTTTCAACGATTTATCTAGCAGTCTGCTACTGCCAGCGGATCATCTCCAAAT 1932
 Qy 542 AspAsnLeuThrValSerThrAsnPhePheMetGlu-----GluLeuLysAlaArg 558
 Db 1933 GACAACTTGAATCCACACTGACACTTTAAATGAAAGTGTTCGTAGATTTAAAGAAAAA 1992
 Qy 559 ThrTyrThrLysValIleLeuMetAspHisValAspTrpLeuAsp-----MetPro 575
 Db 1993 TCATCACTATTCCTATTCATGATCATATGATGATGATGATGATGATGATGATGATGAT 2052
 Qy 576 ValAlaAsnGluLeuAlaGluCysLeuAlaLysGlnValAlaProGlyGlyIleValIle 595
 Db 2053 GCTATTAAATGAATTAATCT---GCTTTGAAAGATGCTTCCGCCAGCGGTAGAGTATTA 2109
 Qy 596 TrpArgSerAlaSerLeuSerProProTyrAlaGluLeuIleGlnLysAlaGlyPhe--- 614
 Db 2110 CTTAGATTCAGCAAGTACAAAGCCCTTGTACTTTGAAACAACTTCAGAACTTTGGGATCCAA 2169
 Qy 615 AspValArgCysIleArgArgAlaThrGlnGlyTyrMetAspArgValAsnMetTyrSer 634
 Db 2170 GAAGAAGAAATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 2229
 Qy 635 Ser 635
 Db 2230 AAT 2232
 RESULT 4
 ID ADW72752
 XX ADW72752 standard; DNA; 2727 BP.
 AC ADW72752;
 XX
 DT 21-APR-2005 (first entry)
 XX

DE Neurospora crassa Btal coding region.
 XX Phospholipid synthesis; transgenic plant; fertilizer; ss; gene;
 KW betaine lipid.
 XX Neurospora crassa.
 OS
 XX Key Location/Qualifiers
 FT CDS 1..2727
 FT /*tag= a
 FT /product= "Btal"
 XX
 PN MO2005009115-A2.
 XX
 PD 03-FEB-2005.
 XX
 PF 15-JUL-2004; 2004WO-US022789.
 XX
 PR 16-JUL-2003; 2003US-00620914.
 XX
 PA (UNMS) UNIV MICHIGAN STATE.
 XX
 PI Benning C, Riekhof W;
 XX
 DR WPI; 2005-112975/12.
 XX P-PSDB; ADW72753.
 PT
 PT New composition comprising a purified DNA having an oligonucleotide
 PT sequence encoding a protein, useful in producing Betaine lipid compounds
 PT e.g., Diacylglycerol-O-4'-(N,N,N-trimethyl) homoserine (DGTS).
 XX
 PS Claim 1; SEQ ID NO 49; 147pp; English.
 XX
 CC The invention relates to a new composition comprising a purified DNA
 CC having an oligonucleotide sequence appearing as ADW72747 (Chlamydomonas
 CC reinhardtii Btal cDNA) or ADW72752 (Neurospora crassa Btal coding region,
 CC encoding Btal proteins which are DGTS (diacylglycerol-O-4'-(N,N,N-trimethyl)homoserine) synthetic enzymes. Btal (and BtaA and BtaB) enzymes
 CC allow the replacement of phospholipids with non-phosphorus containing
 CC lipids in the cell membrane. A transgenic plant expressing the enzymes
 CC would have a reduced need for phosphate-containing fertilizer. Also
 CC included are an RNA transcribed from the purified DNA, antibodies
 CC produced from the protein, a vector comprising the DNA or nucleic acid, a
 CC host cell comprising the vector, transgenic plants comprising the vector,
 CC and a protein translated from the RNA or encoded by the nucleic acid. The
 CC composition is useful in producing Betaine lipid compounds e.g.,
 CC diacylglycerol-O-4'-(N,N,N-trimethyl) homoserine (DGTS). The composition
 CC of the invention reduces the amount of phosphate fertilizer needed for
 CC the optimal growth of crop plants. The present sequence encodes a betaine
 CC lipid synthetic enzyme.
 XX
 SQ Sequence 2727 BP; 616 A; 766 C; 713 G; 632 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 3.73e-81 Length: 2727
 Score: 850.50 Matches: 239
 Percent Similarity: 44.4% Conservative: 109
 Best Local Similarity: 30.5% Mismatches: 240
 Query Match: 24.6% Indels: 198
 DB: 14 Gaps: 23
 US-10-620-914-45 (1-648) x ADW72752 (1-2727)
 Qy 19 LeuGluLysLeuLysLeuSerSerMetLysAspLeuThrValLeuArgHisMet--- 37
 Db 449 TTGGAGAGCTGGAGATG---GGCTTGAANAAGACAAGGAATGGAGACATGTGTAA 505
 Qy 38 ---TrpPheGlySerLysGlyAspAspHisAlaAlaArgLeuGluSerPheTyrGly 56
 Db 506 CCGTGGCCGGGACAG-----GGA 523
 Qy 57 ProGlnAlaAlaPheAlaAlaArgLeuAlaGluArgSerAsnLeuIleTrpValAsp 76

524	CCG-----	GGACAGGAGGAAACCG-ATATCGGTAGAT	555
77	LeuGlyGlyThrGlyGluAsnValAspMetMetAlaAspTyrIleAspLeuAlaLys	96	
556	GTCCGTGGGGCACAGGCTGGAATATCGAAGCCATGCCAAGTTGTCAACGTCCTGAA	615	
97	---PheLysSerIleTyrValValAspLeuCyHisSerLeuCysGluValAlaLysLys	115	
616	TTCTTCAAGACTGTTTACCTAGTAGGACTTTTCGCCCGTCACTTTGTGAAGTGGCTAGGAAG	675	
116	LysAlaLysAlaLysGlyTyrLysAsnValGlnValValGluAlaAspAlaCysGlnPhe	135	
676	CGGTTTCCAGGCTGGGTGGAGATGTGAGAGTTATCTGCACGGATGCTCGCAAGTTT	735	
136	Ala-----ProProGluGlyThr-----	141	
736	AGGCTTGAGGATATGAGGATGTTGACGAGGAGAGTCTGGCTCGGAGATTCTTCGCCT	795	
142	-----AlaThrLeuIle	145	
796	TCCTTGTCCGTTGGTGGGGGAGACGAAGCCGGACGACATGCGGGAGCTGAGTTGATC	855	
146	ThrPheSerTyrSerLeuThrMetIleProPheHisAsnValIleAspGlnAlaCys	165	
856	ACCATGCTTATAGCCTTTCGATGATCCCGATATTTCGATATPCGATCGCTCGAG	915	
166	SerTyrLeuSerGlnAspGlyLeuValGlyValAlaAspPheTyrValSerGlyLysTyr	185	
916	TCTCTGTAGCACCTCAGCGCTTGATTCGGTCTGTGACATTTTACGCCCAGTCGAAAGTC	975	
186	Asp-----LeuProLeuArgGlnMetProTyrSerArg	196	
976	GACTTCACATTCGCAACTACCGGGTGTCTTATGAACCGACACGTTGGCTATTTCGCG	1035	
197	ArgPhePheTyrArgSerIlePheAspIleAspAsnIleAspIleGlyProGluArgArg	216	
1036	CGGAATCTTGGCGCTCGTGGTTCGATGCTGACAGGGTGTCTCTTGAGCCAGCTCGTCGA	1095	
217	AlaTyrLeuGluGlnLysLeuGluArgValTrp-----GluGlnAsnThrGln	232	
1096	GATTATCTCGAGTACAGGTCGGGACTGTCTGTACCGTCAACGCCCGCAACACACTTGT	1155	
233	GlySerIleProTyrValProTyrLeuArgAlaPro-----TyrTyrVal---	247	
1156	GGAGCAATCTTACTACATCTGTTGGG-ATGCCTCAAGAGCCCTTTTCTACGTCGAG	1214	
248	-----TrpIleGlyArgLeuPro-----	253	
1215	TCTACCACAGAAATTGTGGAACACATCGATGCTATTTCGACAGAGTCCCAAGATCATC	1274	
253	-----	253	
1275	ACCCGCTCTAGTGGGCAAAACATTCTCTCTAGCAACAATCGCGTACGCTTTCGAGTCGG	1334	
254	-----SerValGlyHisAlaLeu-----HisGluGluArgValGlu	265	
1335	CCGCACAGCGCGGAGATGCGCTCAAGAGCCCTTCAATACGCCCATCGAACAATCTCGGC	1394	
266	ArgPro-ProMet-----	269	
1395	CAACCTACTCTCCGTCCTTCTTACCAAAATCACCACTGGAGGATCTACTACGACGA	1454	
270	-----PheProProThrPheLeuTyrThrGlnSerTrpGluAs	282	
1455	TCAATCCCGAAGCACACCCAGTTTCAATGACGAGTACATCTACCCCTTTTACCTGGGAAGA	1514	
282	pProGluProAspMetGluValMetGluIleAsnProLysAspThrValLeuThrLeuTh	302	
1515	CTCGGCGTGCACAGAAATCTCTTAACTCGGCCCGCAGACGCTGCTCTAGCCATCAC	1574	
302	rSerGlyGlyCysAsnAlaLeuAsnLeuLeuValGlnGlyAlaGlyGlnValValSerVa	322	
1575	CAGGCGCGGCGACAACTTCTTCTACTCTCATGACAGTCCCGCTCGCGTCACGCAT	1634	

Qy	322	lAspCysAsnProAlaGlnSerAlaLeuLeuGluLeuLysValaIleGlnGlnIle	342
Db	1635	CGACCTAAACCCAGCCAAACACACCTGCTTGAACCTCAAGTCGCCTCTTTTACGACTCT	1694
Qy	342	uGluPheGluAspValTgPlnLeuPheGlyGluGlyValHisProArgIleGluGluLeu	362
Db	1695	GGATTACCCCGAGCTGTGAAGATCTTCGGTGAAGGCAACACCCGACCTTCGCTCACT	1754
Qy	362	uTyfGluLysLeuAlaProPheLeuSerGlnThrSerHisAsnPheTrpSerLysAr	382
Db	1755	GCTCATCTCCAAACTCTCCCTCACCTCTCCGCGCGCGGCTTTCAATACTATGGGTATCCAA	1814
Qy	382	gLeuTrpTyfPheGln-----HisGlyLeuTyfTyfGlnGlyGlyMetGlyLy	398
Db	1815	TGGCGACATATTACCGACCTCGGGGGCGGGTCTCTATGATACCGCGGGCTCCCGATA	1874
Qy	398	sLeuCysTrpValLeuGlnCysLeuAlaValLeuGlyLeuGlyLysThrValLysAr	418
Db	1875	CGCTATCCGGTTCTTCGGCTGGATTTCACACACTCTTCTTCTGCGCTCCGCGCTCGGTG	1934
Qy	418	gLeuAlaAsnAlaProThrMetGlnGluGlnArgArgLeuTrpAspSerAsnMetLeuIl	438
Db	1935	ACTTCTCTACTACCCACCTCGAAGGGCAACGTTCCATCTACCAACCAAGATT-----	1989
Qy	438	eHisPheValLysAsnGlyProLysProLeuValTrpLeuPheValLysPheValSer--	457
Db	1990	-----CGTCCCTGT-----CTGCTCAACCGCTTCGTCAACGG	2021
Qy	458	-LeuValLeuPheAsnLysAlaValLeuTrpPheGlyGlyGlyValProGlyLysGlnTy	477
Db	2022	CCTGGTCTCAGCTCCGACGCTTCTCTGGTCGGCTTGGGGGTGCCCAAGAATCAAAT	2081
Qy	477	rAlaLeuLysAlaAsp-----	483
Db	2082	GGCTATGATGAAGCCGACTACCAACGCGCTTCTATCTCTCTCCACACCCCGCAGCAG	2141
Qy	484	-----GlyIleProIleGluAsnTyfIleAlaArgThrMetAspGl	497
Db	2142	CAAAAGAAAACCAAGCGCGCGCAAGCAATCTCTCCATACACACCTCCACCTTGATCC	2201
Qy	497	yValAlaGluAsnSerHisValArgLysGlnAsnTyfPheTyfTyfAsnCysLeuThrGl	517
Db	2202	CGTTCTCTCCACCTCCACCTTCGCTCGGACACACCTTACTACTCTGCTGTCTCTGGG	2261
Qy	517	yLysPheLeuArgAspAsnCysProThrTyfLeuArgGluAlaAlaPheAlaThrLeuLy	537
Db	2262	ACATACACAGCGCAGTGCCATCCGATTTACTTTCCCTCGCGCCACTCTATCTACG	2321
Qy	537	sSer---GlyValValAspAsnLeuThrValSerThrAsnPhePheMetGluGluLeu--	555
Db	2322	CGCTCTCGAGCGCTTTCAGCGCTTACGCATCCACACCGATGAATAACAGGAGGTGTGGC	2381
Qy	556	-----LysAlaArgThrTyfThrLysValIleLeuMetAspHisValAspTrpLeuAs	573
Db	2382	TAGGTTTCACCGGGTACTTTGACAGTAGCGGTGTGATGATGATGATGATGATGATGATG	2441
Qy	573	pMetPro-----ValAlaAsnGlnLeuAlaGluCy	583
Db	2442	TCCGCTTCGCTGAGGAGGAAAGGAGGCGAGCGAGGAGCAAGTAGGAG	2501
Qy	583	sLeuAlaLysGlnValAlaProGlyGlyIleValIleTrpArgSerAlaSerLeuSerPr	603
Db	2502	GTTGAATCGGCGGTGAAGGTGGGTGAAGAGGTGTTCTTTCAGGAGCGCGGAGTGGAGCC	2561
Qy	603	oProTyfAlaGluLeuIleGlnLysAlaGlyPheAspValArgCysIle-----	619
Db	2562	GTGGTATGTAGGGTTTTTGTGGAGGAAGGGTTTGGACGAGGAGGGTGGTGTCTCGTA	2621
Qy	620	----ArgArgAlaThrGlnGlyTyfMetAspArgValAsnMetTyfSerSerPheTyfMe	638
Db	2622	ATCCGGAAGGGGACGAGGAGTGTATTTGACAGGCTGAATATGATGATGATGATGATGATG	2681


```
Db 1252 TATGCTTCACTGCGGAGATCCTCGTGTGGATCATCTCTTCCGACATCAAGCGGAT 1311
Qy 296 AspThrValLeuThrLeuThrSerGlyGlyCysAsnAlaLeuAsnLeuLeuValGlnGly 315
Db 1312 GATGTTATCTTGGCCATTTACAGCGCGGAGACAATATTTGGATTACCTTCAGAAAGAT 1371
Qy 316 AlaGlyGlnValValSerValAspCysAsnProAlaGlnSerAlaLeuLeuGluLeuLys 335
Db 1372 CCAGCAGAGTTTCATGCGAGTCTGATGCGGATGCTCTGGAAGATCTTTGGCGAGGAAAA 1491
Qy 336 LysValAlaIleGlnGlnLeuGluPheGluAspValTrpGlnLeuPheGlyGluGlyVal 355
Db 1432 GTTGCTAGTTTCATGCGCCCTTGTGTCGCGATGCTCTGGAAGATCTTTGGCGAGGAAAA 1491
Qy 356 HisProArgIleGluGluLeuLeuThrGlyGluLeuAlaProPheLeuSerGlnThrSer 375
Db 1492 CAGCCAGAAATTCAGGGAACCTCTCTTCCTGCTCAGCGCTCACCTCTCCAGCCAGGCA 1551
Qy 376 HisAsnPheTrpSerLysArgLeuTrpTyrPheGln-----HisGlyLeuTyr 391
Db 1552 TTCAGTACTGGCTTGAGCACACTCACATTTTCACTTCAAAATATGGCAAGGACTTTAT 1611
Qy 392 TyrGlnGlyMetGlyLysLeuCysTrpValLeuGlnCysLeuAlaValValLeuGly 411
Db 1612 GAAACCGTGGCTCGGCCAGCCATCAAGATGTTGCGTACCTGTTCAAGGTGTTTGGC 1671
Qy 412 LeuGlyLysThrValLysArgLeuAlaAsnAlaProThrMetGluGlnGlnArgArgLeu 431
Db 1672 CTTGAGGGTCAAGTGAAGAGCTCTGCGAGCGCAGACTCTTGTGAGCAACGTGAGATC 1731
Qy 432 TrpAspSerAsnMetLeuIleHisPheValLysAsnGlyProLysProLeuValTrpLeu 451
Db 1732 TGGCCAAAGATTGCGCGCTACTCATG-----AGCAAGCCCTTCAATGG--- 1776
Qy 452 PheValLysPheValSerLeuValLeuPheAsnLysAlaValLeuTrpPheGlyGly 471
Db 1777 -----GCGTGTGTCAGCAGCAGATGTTGCGGTGGAAGCGCGCGGC 1818
Qy 472 ValProGlyLysGlnTyrAlaLeuLysAlaAsp----- 483
Db 1819 GTGCTCTGAACACAGCGGAATATGATGTTGAGCACTACTTCAAGAGACTGGGCTGACC 1878
Qy 484 -----GlyIleProIleGluAsnTyrIleAlaArg 493
Db 1879 AAGACATGAAACAGCGCAGGATATCAGTGGCGGTCGATCTGCGAATATGTTGGAC 1938
Qy 494 ThrMetAspGlyValAlaGluAsnSerHisValArgLysGlnAsnTyrPheTyrAsn 513
Db 1939 AACTGGACCCGTCGTTAAGGAAACCATGATGATCAGCAACGATACTACTTTTACTTCTG 1998
Qy 514 CysLeuThrGlyLysPhe-----LeuArgAsp 522
Db 1999 TGTCTCAGGGGCGAGTTTCAAGACGAGTGCAAAACAAACCAATCAAGTGATCCGAGAT 2058
Qy 523 AsnCysProThrTyrLeuArgGluAlaAlaPheAlaThrLeuLysSer---GlyValVal 541
Db 2059 GCCAC-CCACATACTTGTCTCTCAGCCCATGATGAGCTCTTTCGCCGGTGCAATT 2117
Qy 542 AspAsnLeuThrValSerThrAsnPhePheMetGluLeuLys-----AlaArg 558
Db 2118 GACGGCTACGTAATCACACAGATGAGATCAACGAGTGAATCAAGCAATCACTTCCCGCA 2177
Qy 559 ThrTyrThrLysValIleLeuMetAspHisValAspTrpLeuAspMetProValAlaAsn 578
Db 2178 AGCTGCAATTCGCTGTGATCATGACTCGATGCTGTTTCGAT---CCGGAAGGCACA 2234
Qy 579 GluLeu-----AlaGluCysLeuAlaLysGlnValAlaProGlyGlyValIle 595
Db 2235 GAACATCGATCCAGGCCAGAACTCAACACGCCCTGAAATGGAGCGCGCATTTCTC 2294
Qy 596 TrpArgSerAlaSerLeuSerProProTyrAlaGluLeuIleGlnLysAlaGlyPheAsp 615
Db 2295 CTCCGCTCTGCAGCATCGAGCCCTTGGTATATCAAGCAGATTCGAAGAGAAATGGGTTTCA 2354
```

```
Qy 616 ValArgCysIle---ArgArgAlaThrGlnGlyTyrMetAspArgValAsnMetTyrSer 634
Db 2355 GCCCGCGGTGCGGCCCGCTCCCGGCTCTTGCATGCATCGTGTGACATGATATGCA 2414
Qy 635 SerPheTyrMetAlaArgArg 641
Db 2415 TCGACGTGGATCTGTACCAAG 2435
```

RESULT 6

ADW72751

ID ADW72751 standard; DNA; 3427 BP.

XX ADW72751;

XX 21-APR-2005 (first entry)

XX Neurospora crassa Btal gene.

XX Phospholipid synthesis; transgenic plant; fertilizer; ds; gene;

XX betaine lipid.

XX Neurospora crassa.

XX WO2005009115-A2.

XX 03-FEB-2005.

XX 15-JUL-2004; 2004WO-US022789.

XX 16-JUL-2003; 2003US-00620914.

XX (UNMS) UNIV MICHIGAN STATE.

XX Benning C, Riekhof W;

XX WPI; 2005-112975/12.

XX P-PSDB; ADW72753.

XX New composition comprising a purified DNA having an oligonucleotide

XX sequence encoding a protein, useful in producing Betaine lipid compounds

XX e.g., Diacylglycerol-O'-4'-(N,N,N-trimethyl) homoserine (DGTS).

XX Example 5; SEQ ID NO 48; 147pp; English.

XX The invention relates to a new composition comprising a purified DNA

XX having an oligonucleotide sequence appearing as ADW72747 (Chlamydomonas

XX reinhardtii Btal cDNA) or ADW72752 (Neurospora crassa Btal coding region,

XX encoding Btal proteins which are DGTS (diacylglycerol-O'-4'-(N,N,N-

XX trimethyl)homoserine) synthetic enzymes. Btal (and BtaA and BtaB) enzymes

XX allow the replacement of phospholipids with non-phosphorus containing

XX lipids in the cell membrane. A transgenic plant expressing the enzymes

XX would have a reduced need for phosphate-containing fertilizer. Also

XX included are an RNA transcribed from the purified DNA, antibodies

XX produced from the protein, a vector comprising the DNA or nucleic acid, a

XX host cell comprising the vector, transgenic plants comprising the vector,

XX and a protein translated from the RNA or encoded by the nucleic acid. The

XX composition is useful in producing Betaine lipid compounds e.g.,

XX Diacylglycerol-O'-4'-(N,N,N-trimethyl) homoserine (DGTS). The composition

XX of the invention reduces the amount of phosphate fertilizer needed for

XX the optimal growth of crop plants. The present sequence encodes a betaine

XX CC lipid synthetic enzyme.

XX SQ Sequence 3427 BP; 805 A; 938 C; 867 G; 814 T; 0 U; 3 Other;

Alignment Scores:

Pred. No.:	5.92e-76	Length:	3427
Score:	804.00	Matches:	238
Percent Similarity:	43.2%	Conservative:	112
Best Local Similarity:	29.4%	Mismatches:	256
Query Match:	23.2%	Indels:	205
DB:	14	Gaps:	24

Db 1299 ATCTCCTTATCTTCTCCCGCAATCACCAAGGAGATGAGAAAGCAATGGAGGAA 1358
Qy 251 ----- 251
Db 1359 CGCTCAAGAAATTCGTTCCAAAGCCTACGAGTCGGCGTTATCAACCTGAGCGGACCT 1418
Qy 252 -----LeuProSerValGlyHisAlaLeuHisGluGluArgVal----- 264
Db 1419 CCGCTTTCCTCATCTTACCAAGATCAACCACTGCCGATCTTCTACCAATGATCTGCT 1478
Qy 265 -GluArgProMetPheProThrPheLeuTyrThrGlnSerTrpGluAspProG1 284
Db 1479 GCCCAAGCAGCAGGTTCAAGATGAATACATTTATGCTTCACTGGGAAGATCTCTCG 1538
Qy 284 uProAspMetGluValMetGluLeuLeuProLysAspThrValLeuThrLeuThrSerG1 304
Db 1539 TGTGATCATCTCTCTGACATCAAGCGGATGATGTTATCTTGGCCATTACCAAGCGC 1598
Qy 304 yGlyCysAsnAlaLeuAsnLeuValGlnGlyAlaGlyGlnValValSerValAspCy 324
Db 1599 CGGAGACATATTTTGGATTACCTTCCAGAGAGTCCACGAGAGTTTCATGAGTCGACCT 1658
Qy 324 sAsnProAlaGlnSerAlaLeuLeuGluLeuLysValAlaLeuGlnLeuGluPh 344
Db 1659 GAATCCTTAACAGATCATTTGTTGAATCAAGTTGCTAGTTTCATGGCCCTTGGTCA 1718
Qy 344 eGluAspValTrpGlnLeuPheGlyGluGlyValHisProArgIleGluGluLeuTyrG1 364
Db 1719 TCGCGATGTTCTGGAAGATCTTTGGCGAGGGAACACACCCAGAAATTCAGGGAATCTCTCAT 1778
Qy 364 uLysLeuAlaProPheLeuSerGlnThrSerHisAsnPheTrpSerLysArgLeuTr 384
Db 1779 TTTCTGCTCAGCGCTCACCTCTCCAGCCAGGATTCAGTACTGCTTGGCAGCACACTCA 1838
Qy 384 pTyrPheGln-----HisGlyLeuTyrTyrGlnGlyGlyMetGlyLysLeuCy 400
Db 1839 CATTTTCACTCAAAATATGGCAAGGACTTATGAACCGTGGCTCGCGCCACGCCAT 1898
Qy 400 sTrpValLeuGlnCysLeuAlaValValLeuGlyLeuGlyLysThrValLysArgLeuAl 420
Db 1899 CAAGATGGTTCCGTACCTGTTCAAGGTGTTGGCTTTGAGGTCAGGTGAAGAAGCTCTG 1958
Qy 420 aAsnAlaProThrMetGluGlnArgArgLeuTrpAspSerAsnMetLeuLeuHisPh 440
Db 1959 CGAGGCGCAGACTCTTGTGAGCAACGTGAGATCTGGCCAAAGATTGCGCGCTGACTCAT 2018
Qy 440 eValLysAsnGlyProLysProLeuValTrpLeuPheValLysPheValSerLeuValLe 460
Db 2019 G-----AGCAAGCCCTTCATTCG-----CGGTTGT 2045
Qy 460 uPheAsnLysAlaValLeuTrpPheGlyGlyGlyValProGlyLysGlnTyrAlaLeu11 480
Db 2046 CAGCAGCGATGTTTCGCGTGAAGGCGCGGGGCTGCTCGAACCAGCGGATATGAT 2105
Qy 480 eLysAlaAsp----- 483
Db 2106 CGTTGACGACTACTTCAAGAGACTGGGCTGACCAAGGACATCAACCCAGGCAAGGATAT 2165
Qy 484 ----GlyIleProIleGluAsnTyrIleAlaArgThrMetAspGlyValAlaGluAsnSe 502
Db 2166 CAGTGGCCGTCGATCTGGCAATATGTTGTGGACACACTGGACCCCGCTGTTAACGAAC 2225
Qy 502 rHisValArgLysGlnAsnTyrPheTyrTyrAsnCysLeuThrGlyLysPheLeuArg-- 521
Db 2226 CATGATCAGCAACGATAACTACTTTTACTTCTCTGTTCTCCAGGGGCGAGTTTTCAGAGC 2285
Qy 522 -----AspAsnCys---Pr 525
Db 2286 GTGAGTTTGTCCCAAGAGTGCACAAACCAATCTAGTGTATCCGAGATGCCACCC 2345
Qy 525 oThrTyrLeuArgGluAlaAlaPheAlaThrLeuLysSer---GlyValValAspAsnLe 544
Db 2346 CACATCTTCTCTCCTCAGGCCCATGTGAAGCTCTCTTTCGCCCGGTCATTTGACGGCCT 2405

Qy 544 uThrValSerThrAsnPhePheMetGluGluLeuLys-----AlaArgThrTyrTrh 561
Db 2406 ACGTATTTACACAGATGAGATCAACGAAGTGATCAAGCGAATCACTCCCGAAGCCTGAC 2465
Qy 561 rLysValIle----- 564
Db 2466 AATCGTGTGGTAGCAGCGCCCTTGTATGATTTTGGCCAAATGACCGATCTAACCCATTG 2525
Qy 565 ----LeuMetAspHisValAspTrpLeuAspMetProValAlaAsnGluLeu----- 580
Db 2526 CCAGATCATGGACTCCGATGGACTGGTTGAT-----CCGGAAGGCACAGACATCGATCCA 2582
Qy 581 -AlaGluCysLeuAlaLysGlnValAlaProGlyGlyIleValIleTrpArgSerAlaSe 600
Db 2583 AGCCAGAAACTCAACACCGCCCTGAAATGGACGCGCGCATTTCTCTCGCTCTGCGAG 2642
Qy 600 rLeuSerProProTyrAlaGluLeuLeuGlnLysAlaGlyPheAspValArg----- 617
Db 2643 CATCGAGCCTTGGTATATATCAAGCAGTTTCAAGAGATGGGTTTCAACAGCCCGCTGTCGG 2702
Qy 618 -----CysIleArgArg----- 621
Db 2703 CGCCCGCTTCCCGGCTCTTGCATCGATCGTAAGTTCTCGATATTCCTCGACTGTTTAC 2762
Qy 622 ----Ala-ThrGlnGlyTyrMetAspArgValAsnMetTyrSerSerPheTyrMetAlaA 640
Db 2763 CAATGCAATCGCTGACGCTAC---GACAGTGTGAACATGTATGATCATCGACGTGATCTGTA 2819
Qy 640 rArg 641
Db 2820 CCAAG 2824
RESULT 8
ID ADR84295
XX ADR84295 standard; DNA; 8929 BP.
AC ADR84295;
XX 04-NOV-2004 (first entry)
DT XX
XX Aspergillus fumigatus essential gene genomic sequence #106.
DE Aspergillus fumigatus essential gene genomic sequence #106.
XX Fungicide; Aspergillus fumigatus infection; Farmer's lung disease;
KW drug screening; ds.
XX Aspergillus fumigatus.
OS
XX WO2004067709-A2.
PN 12-AUG-2004.
XX 16-JAN-2004; 2004WO-US001099.
PF 17-JAN-2003; 2003US-0441281P.
PR 13-JUN-2003; 2003US-0478196P.
XX (ELIT-) ELITRA PHARM INC.
PA (ELIT-) ELITRA CANADA LTD.
XX Jiang B, Hu W, Lemieux S, Roemer T;
PI WPI; 2004-594200/57.
XX P-PSDB; ADR86056.
XX New purified or isolated Aspergillus fumigatus nucleic acid molecule
PT encoding a gene product, useful for diagnosing and/or treating invasive
PT fungal infections, such as Farmer's lung disease.
XX Claim 3; SEQ ID NO 106; 164pp; English.
XX The present invention relates to Aspergillus fumigatus genes that are
XX essential and are potential targets for drug screening. The methods and
CC

compositions of the present invention are useful for diagnosing and/or treating invasive *Aspergillus fumigatus* infection, including the allergic forms of the disease, such as Farmer's lung disease. They can also be used in various drug discovery purposes, such as expression of the recombinant protein, hybridization assay and construction of nucleic acid arrays. The present sequence represents an *Aspergillus fumigatus* essential gene full length genomic sequence, used during diagnosis and drug development in the invention. These genes share a high degree of sequence conservation with known essential genes of *Candida albicans*. The sequence data for this patent is not represented in the printed specification, but was obtained in electronic format from WIPO.

xx SQ Sequence 8929 BP; 2141 A; 2280 C; 2289 G; 2219 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 5.06e-64 Length: 8929
Score: 700.00 Matches: 219
Percent Similarity: 38.0% Conservative: 94
Best Local Similarity: 26.6% Mismatches: 232
Query Match: 20.2% Indels: 278
DB: 13 Gaps: 25

US-10-620-914-45 (1-648) x ADR84295 (1-8929)

QY 73 IleTyrValAsp----- 76
DB 3432 ATCTGGTGTGATTAAGACACCTCGGTATCGCTATTGTTGCGAAGCGCTGACACTCGTT 3491
QY 77 -----LeuGlyGlyThr-GlyGlu----- 83
DB 3492 GCTTAGATTGGCGGAGCAGCTGGTAAAGTGTCAAAATCGAGTCAACGCTTTGCTTCGAGGC 3551
QY 84 -----AsnValAspMetMetAlaAspTyrIleAspLeuAlaLysPhe 97
DB 3552 CTAATCGTAGGATAGCTACAAACATTTAGGCGCATGGCTTCTTTCTACCACTTACCCAAAT 3611
QY 97 eLysSer---IleTyrValValAspLeuCysHisSerLeuCysGluValAlaLysLysLy 116
DB 3612 CTTCCTCGACGTGTACTTGTGACCTGTCTCCCTCTCTTTGCGAGGTTCGCGGTACAGC 3671
QY 116 sAlaLysAlaLysGlyTrpLysAsnValGlnValValGluAlaAspAlaCysGlnPhe-- 135
DB 3672 ATTCGAGCGCTAGCTGGTGAAGATGTCAAGTGTCTGTCAAGATCGCGCTCTTTCCG 3731
QY 136 -----AlaProGluGlyThrAlaThrLe 144
DB 3732 TCTCCCCCATGAGAAAGTCGACCCCGTGGCCAGCTCCATCAACTGCGAGCGCAGATCT 3791
QY 144 uIleThrPheSerTyrSerLeuThrMetIlePro----- 155
DB 3792 TATCACCATGAGCTACAGCTTGTCATGATTCCTGTTAGTTTGGATTAAACAAGCATT 3851
QY 156 -----ProPheHisAsnValIleAspGlnAlaCysSerTy 167
DB 3852 GAGGCGCGCTACTAAGTGAATACAGACTACTACAGTGTCTGCTGACCTCCCTGACCGCCT 3911
QY 167 rLeuSerGlnAspGlyLeuValGlyValAlaAspPhe----- 179
DB 3912 CTTCGAGCGCTCGGTATCTTGTGTCTGCGACTTTTATGTTAAGTAGCTAGCCTACAGCCTT 3971
QY 179 ----- 179
DB 3972 CTGACGTGATAGATACTAATTTATCGTATAGTGCMAAGTATCGTGGACGTTTCGCGCG 4031
QY 180 ----TyrValSerGlyLysTyrAspLeuProLeuArgGlnMetProTyrSerArgArgph 198
DB 4032 CAATACATTGGTGTGCTTTCAAT-----CGCCACGCTTAACCTGGCTGGGTCTGTC 4082
QY 198 ePheTyrArgSerIlePheAspIleAspAsnIleAspIleGlyProGluArgArgAlaTy 218
DB 4083 CTTTTGGCGGCGCTGGTTCGATGGCGGCGGTGTGAGTGTGGAAGCAGCAGCAGACTA 4142
QY 218 rLeuGluGlnLysLeuGluArgValTrpGluGlnAsnThrGln-----Glyse 234

DB 4143 TCTGGAATACCGCTTTGGAAACCGTCACTCTCTGCAGCAGCGCAATATCTCTTCGCGCG 4202
QY 234 rIleProTyrValProTyrLeuArgAlaProTyrTyrValTrpIleGlyArg----- 251
DB 4203 CATT-----CCTTACTACATCTTCGTTGGACGTCACAAAGAA 4238
QY 251 ----- 251
DB 4239 TCTTGATCAATCTCTCGGTTCAGGACAAATTGAAATGTTGGACGCTCTCTTACTGA 4298
QY 251 ----- 251
DB 4299 ATCTCTTATCTTCTCCGCAATCACACAGGAGATGGAGAAAGCAATGGAGAGAA 4358
QY 251 ----- 251
DB 4359 CGCTCAAGAAATTCGTTCCAAAGCTACGAGTCGCGCGTTATCACTGAGCGCGAACCT 4418
QY 252 ----LeuProSerValGlyHisAlaLeuHisGluGluArgVal----- 264
DB 4419 CCGCTTCCCTCATCTCTTACCAGATCACCACTGCGCATCTTCTACAATGATCTGCT 4478
QY 265 -GluArgProProMetPheProThrPheLeuTyrThrGlnSerTyrGluAspProG1 284
DB 4479 GCCAAGCACACGACGCTCAAGAAATGAATACATTTATGCTTCAACTGGGAAGATCCTCG 4538
QY 284 uProAspMetGluValMetGluIleAsnProLysAspThrValLeuThrLeuThrSerG1 304
DB 4539 TGTGGATCATCGTCTTCTCGACATCAAGCGGATGATGTTATCTTGGCCATACACAGCG 4598
QY 304 yGlyCysAsnAlaLeuAsnLeuValGlnGlyAlaGlyGlnValValSerValAspCy 324
DB 4599 CGAGACAAATATTTGGATTACCTTCAGAAGAGTCCACGAGAGTTTCATCGACTCGACCT 4658
QY 324 sAsnProAlaGlnSerAlaLeuLeuGluLeuLysLysValAlaIleGlnGlnLeuGluPh 344
DB 4659 GAATCTCAACAGAAATCAATTTGCTGAACCTCAAGTTGTAGTTTCATGCGCCCTTGCTCA 4718
QY 344 eGluAspValTrpGlnLeuPheGlyGluGlyValHisProArgIleGluGluLeuTyrG1 364
DB 4719 TCCTGATGTCTGAAGATCTTTCGCGAGGAAACCCCAAGAAATTCAGGAACTCTCAT 4778
QY 364 uLysLysLeuAlaProPheLeuSerGlnThrSerHisAsnPheTrpSerLysArgLeuTr 384
DB 4779 TTCTGCTCAGCGCTCACCTCTCCAGCAGGCAATCCAGTACTGCTGCTGAGCACACTCA 4838
QY 384 pTyrPheGln-----HisGlyLeuTyrTyrGlnGlyGlyMetGlyLysLeuCy 400
DB 4839 CATTTTCACTTCAAAATATATGCAAAAGACTTTATGAACCGTGGCTCGCGCCACGCCAT 4898
QY 400 sTrpValLeuGlnCysLeuAlaValValLeuGlyLeuGlyLysThrValLysArgLeuAl 420
DB 4899 CAAGATGTTTCGTTACCTGTTCAAGGTGTTGCTTGGCTTGGAGGTTCAGTGAAGAGCTCTG 4958
QY 420 aAsnAlaProThrMetGluGlnArgArgLeuTrpAspSerAsnMetLeuLeHisPh 440
DB 4959 CGAGCGCAGACTCTTCTGAGCAACGTGAGATCTGGCCAAAGATTTCGCGCGTACTCAT 5018
QY 440 eValLysAsnGlyProLysProLeuValTrpLeuPheValLysPheValSerLeuValLe 460
DB 5019 G-----AGCAAGCCCTTCAATGG-----CGGTGTGT 5045
QY 460 uPheAsnLysAlaValLeuTrpPheGlyGlyValProGlyLysGlnTyrAlaLeuIl 480
DB 5046 CAGCACCAATGTTTCGTTGGAAGCGCGCGGCTGCTCGAAACACCGCGGAATATGAT 5105
QY 480 eLysAlaAsp----- 483
DB 5106 CGTTGACGACTACTTCAAGAGACTGGGCGCTGACCAAGACATGAACCCAGCGCAAGGATAT 5165
QY 484 ----GlyIleProIleGluAsnTyrIleAlaArgThrMetAspGlyValAlaGluAsnSe 502

Db 5166 CAGTGGCCGTCGATCTGGCAATATGTTGGACACACTGGACCCCTCTTAAACGAAC 5225
 Qy 502 rHisValArgLysGlnAsnTyrPheTyrAsnCysLeuThrGlyLysPheLeuArg-- 521
 Db 5226 CATGATCAGCAACGATACACTTTTACTTCTCTCCAGGGGAGTTTTCNAGACG 5285
 Qy 522 -----AspAsnCys-----Pr 525
 Db 5286 GTGAGTTTGTCCCAAGAGTGCAAAACCAATCTAAGTGTATCCGAGATGCCACC 5345
 Qy 525 oThrTyrLeuArgGluAlaAlaPheAlaThrLeuLysSer---GlyValValAspAsnLe 544
 Db 5346 CACATCTTGTCTCTCCAGGCCCATGTGAAGCTCTTCCCGCGGTCATTTGACGGCCT 5405
 Qy 544 uThrValSerThrAsnPhePheMetGluGluLeuLys-----AlaArgThrTyrTh 561
 Db 5406 ACGTATTTCACACAGATGATCAACGAAGTGATCAAGCAATCACTCCCGAAGCTGAC 5465
 Qy 561 rLysValLe-----564
 Db 5466 AATCGCTGTGTGTCGACGCCCTTGTATGATTTGCCCAATGACCGATCTAACCCATTG 5525
 Qy 565 ----LeuMetAspHisValAspTyrLeuAspMetProValAlaAsnGluLeu----- 580
 Db 5526 CCAGATCATGACTCGATGACTGTTTCGAT---CCGAAAGGCACAGAAGCATCGATCCA 5582
 Qy 581 -AlaGluCysLeuAlaLysGlnValAlaProGlyGlyLeuValIleTyrArgSerAlaSe 600
 Db 5583 AGCCAGAACTCAACACGCCCTGAAATGAGCGCGCATCTCTCTCGCTCTGCCAG 5642
 Qy 600 rLeuSerProTyrAlaGluLeuIleGlnLysAlaGlyPheAspValArg----- 617
 Db 5643 CATCGAGCCTTGTATATCAAGCATTCGAAGATGAGTGTTCACAGCCCGGTGTGCG 5702
 Qy 618 -----CysIleArgArg-----621
 Db 5703 CGCCCGCTCCCGCGCTCTTGCATCGATCGGTAAAGTTCGATATCCCTGATGTTTAC 5762
 Qy 622 ----Ala-ThrGlnGlyTyrMetAspArgValAsnMetTyrSerSerPheTyrMetAlaA 640
 Db 5763 CAATGCATCGCTGACGTAC---GACAGTGTGAACATGATGATCATCGACGTGATCTGTA 5819
 Qy 640 rArg 641
 Db 5820 CCAAG 5824
 RESULT 9
 ABS58051 standard; DNA; 1252 BP.
 AC ABS58051;
 XX 05-FEB-2003 (first entry)
 DE Gene encoding Rhodobacter sphaeroides btaA protein.
 XX Betaine lipid production; betaine lipid compound; DGTS;
 KW diacylglycerol-0-4'-(N,N,N-trimethyl) homoserine; agricultural industry;
 KW phosphate-containing fertiliser; transgenic; btaA; gene; ds.
 XX Rhodobacter sphaeroides.
 XX Key Location/Qualifiers
 CD 1. .1251
 FT /*tag= a
 FT /partial
 FT /product= "BtaA protein"
 FT /transl_except= (pos:1..3, aa:Met)
 FT /note= "This sequence lacks a start codon"
 XX W020283844-A2.
 PN 24-OCT-2002.

XX 09-APR-2002; 2002WO-US011134.
 PF 13-APR-2001; 2001US-0283812P.
 PR 08-APR-2002; 2002US-00283812.
 XX (UNMS) UNIV MICHIGAN STATE.
 XX Benning C, Riekhof W, Klug R;
 XX WPI; 2003-058632/05.
 DR P-PSDB; ABG721176.
 XX New composition comprising an isolated and purified DNA molecule, useful
 PT for producing Betaine lipids, e.g. Diacylglycerol-0-4'-(N,N,N-trimethyl)
 PT homoserine (DGTS) for agricultural applications.
 XX Claim 1; Fig 7; 109pp; English.
 XX The present invention relates to compositions and methods for producing
 CC betaine lipids. The method comprises the expression of recombinant
 CC enzymes (e.g. from Rhodobacter sphaeroides) in host cells such as
 CC bacteria, yeast, and plants to produce betaine lipid compounds including
 CC diacylglycerol-0-4'-(N,N,N-trimethyl) homoserine (DGTS). The methods and
 CC compositions of the invention are useful in agricultural applications,
 CC such that the amount of phosphate-containing fertiliser required for the
 CC growth of a particular plant is decreased. The polynucleotide sequences
 CC encoding the recombinant enzymes may be used to produce vectors which can
 CC be used to produce transgenic plants. The present sequence encodes R.
 CC sphaeroides btaA protein
 XX SQ Sequence 1252 BP; 152 A; 488 C; 424 G; 188 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 1.5e-30 Length: 1252
 Score: 379.50 Matches: 127
 Percent Similarity: 43.1% Conservative: 68
 Best Local Similarity: 28.1% Mismatches: 182
 Query Match: 11.0% Indels: 75
 DB: 8 Gaps: 20
 US-10-620-914-45 (1-648) x ABS58051 (1-1252)
 Qy 231 ThrGlnGlySerIleProTyrValProTyrLeuArgAlaProTyrTyrValTrpIleGly 250
 Db 4 ACGCAGTTCGCTCCACCCACCTGCC-----GCCCGCCG-----GTTGCC 45
 Qy 251 ArgLeuProSerValGlyHisAlaLeuHis-----GluGluArg 263
 Db 46 CGC-----CAGATCGCGCGCGCGTGCACCGCACGCTGCTTCTCAGCGCGGAGGACTG 99
 Qy 264 ValGluArgProMetPheProThrPhe-----LeuTyrThrGlnSerTyr 280
 Db 100 ATGGAGCG-----ATGTTCTCGCGCTCTTCCAGGGCTCGTCTATCCGAGATCTGG 153
 Qy 281 GluAspProGluProAspMetGluValMetGluIleAsnProLysAspThrValLeuThr 300
 Db 154 GAGATCCGCGGTGACATGCGCGCCCTCGCCATCGCCCGCGGACCGGCTGGTGCC 213
 Qy 301 LeuThrSerGlyCysAsnAlaLeuAsnLeuValGlnGlyAlaGlyGlnValVal 320
 Db 214 ATCGCTCGGGGGTTCGCAACGTGCTTCTCTATCTCAGCAGGGGCGGCTCGATCTC 273
 Qy 321 SerValAspCysAsnProAlaGlnSerAlaLeuGluLeuLysValAlaIleGln 340
 Db 274 GCGGTGATCTCTCGCCCGCCCATGTGGCGCTGGGCGGCTGAAGCTCGCGCGCGCGG 333
 Qy 341 GlnLeu---GluPheGluAspValTyrGlnLeuPheGlyGluGlyValHisProArgIle 359
 Db 334 ACGCTGCCGACCATCGCGCTTCTTCTGATCTCTTCTCGGTGCGCGCAGACCTCGCGCAAT 393
 Qy 360 GluGluLeuTyrGluLysLysLeuAlaProPheLeuSerGlnThrSerHisAsnPheTyr 379

Score: 329.00 Matches: 118
Percent Similarity: 42.8% Conservative: 66
Best Local Similarity: 27.4% Mismatches: 184
Query Match: 9.5% Indels: 63
DB: 14 Gaps: 15

US-10-620-914-45 (1-648) x ADW72725 (1-1248)

QY 248 TrpIleGlyArgLeuProSer-----ValGlyHisAlaLeuHisGluGluArgValGlu 265
Db TGAAGAGCGCGTACAGAACCGCGCGCTTCCAAAGCGCGCATCTCCGAGCGGTGT 107
QY 266 ArgProMetPheProPro-ThrPheLeuTyThrGlnSerTrpGluAspProGluPr 285
Db CGCTTCCTCTTTCGCGCTCGTC-----TATCCGAGATCTGGAGAACCCCGATGT 161
QY 285 oAspMetGluValMetGluIleAsnProLysAspThrValLeuThrLeuThrSerGlyG 305
Db CGACATGGAGCGCATGAGCTTGGTCAGGCGCATCGCATCGTCACAACTCGCTTCGCGCG 221
QY 305 YCysAsnAlaLeuAsnLeuLeuValGlnGlyAlaGlyGlnValValSerValAspCysAs 325
Db CTGCAACATCTCGCTTACCTACCGTTCGCGCGCAGCGATCGACGCGTGCACCTCAA 281
QY 325 nProAlaGlnSerAlaLeuLeuGluLeuLysLysValAlaIleGlnGlnLeuGluPheG 345
Db CGCGCGCCACATCGCGTGAACCGCATGAAGCTGGAGCGCGTGCCTGCTGCGCTCGCA 341
QY 345 u---AspValTrpGlnLeuPheGlyGluGlyValHisProArgIleGluGluLeuTyG 364
Db GGGCGATCTGTTCCGCTTTTCGCGCGCGCGCACAGCCACCAATTCGACGCTATGA 401
QY 364 uLysLysLeuAlaProPheLeuSerGlnThrSerHisAsnPheTrpSerLysArgLeuTr 384
Db CGCTTTATTGCGCGCATCTCGATCGGTGAGCGCGCACTATTGGAGCGCGCACTG 461
QY 384 p-----TyrPheGlnHisGlyLeuTyThrGlnGlyMetG 397
Db GCGTGTGCGCGCGCATCGCGCTTCGACCGCAATTTCTACACAGCGCGCTGCTCGG 521
QY 397 YLysLeuCysTrpValLeuGlnCysLeuAlaValValLeuGly----- 411
Db CTGTTTCATCGCGCATGGCGCATCGCGCGCGCATCGCGCGCGCAATTT 557
QY 412 -LeuGlyLysThrValLysArgLeuAlaAsnAlaProThrMetGluGluGlnArgArgLe 431
Db CTTTCGCGCTCAACCGCGCCACATGATGGAAGCCAGCAATATCGCGGAGCGCGCTT 617
QY 431 uTrpAspSerAsnMetLeuIleHisPheValLysAsnGlyProLysProLeuValTrpLe 451
Db CTTCAACGAGGAGCTGGCGCGCTCTTCGACAAAG-----AAGCTTTTGAATGGGC 668
QY 451 uPheValLysPheValSerLeuValLeuPheAsnLysAlaValLeuTrpPheGlyGly 471
Db GACCTCGGTAAGCGCTCGCTG-----TTCCGCTCGG 701
QY 471 YValProGlyLysGlnTyr-----AlaLeuIleLysAlaAspGlyLeProIleGluAs 489
Db CTTTCGCGCGCGCAGTACGATTCCTGATCCTCAGGCGCGCGCACCATCGCGCGCT 761
QY 489 nTyrlleAlaArg-----ThrMetAspGlyValAlaGluAsnSerHisValAr 505
Db TCTGAAGGCGCGCTGGAAGAGCTCGCTGCGATTTTCCCTCGAAAC----- 810
QY 505 gLysGlnAsnTyThrPheTyrAsnCysLeuThrGlyLysPheLeuArgAspAsn----- 523
Db AATATTTCGCTGGCAGCGCTTTGCGCGCGCTATTCCAATCCCGGTGAGGC 863
QY 524 ----CysProThrTyrlleArgGluAlaAlaPheAlaThrLeuLysSerGlyValValAs 542
Db CGCGCTCGCGCTATCTGGAAGAGCAGACATGACGAACCATCCGC-----GGCAATATCGA 920
QY 542 pAsnLeuThrValSerThrAsnPhePheMetGluGluLeu-----LysAlaArgTh 559

Db 921 CGCGTCCGCATCCACCATCCCAATCTGATCGAATTCCTCGCGCAAGCAGCGGCAC 980
QY 559 rTyThrLysValIleLeuMetAspHisValAspTrpLeuAspMetProValAlaAsnG 579
Db 981 CGTCGATCGCTTCATCTCTCGATCGCGAGGACTGATGACCGATGACCGATCAACGC 1040
QY 579 uLeuAlaGluCysLeuAlaLysGlnValAlaProGlyGlyLeValIleTrpArgSerAl 599
Db 1041 GCTGTGTCGGAATCAGCGCACCGCTCCGCGAGCGCGCGCTCATCTTCGACCGC 1100
QY 599 a-----SerLeuSerProProTyrrAlaGluLeuIleGlnLysAlaGlyPheAsp-V 616
Db 1101 CGCGAGCGCGCTCTCGCGCGCGCTCTCGATCTCGCTCTCGACCGATGGGACTA 1160
QY 616 aArgCysIleArgArgAlaThrGlnGlyTyrrMetAspArgValAsnMetTyrrSerP 636
Db 1161 TCAGGACGAGCGCTCGCGGAATTCGCGCACGC-GACCGTTCGGCCATCTATGGCGCT 1219
QY 636 heTyrrMetAlaArgArgLysGlyAla 644
Db 1220 TCACCTCTATGTGAAGCGCACGGCA 1245

RESULT 13
ABS58082
ID ABS58082 standard; DNA; 1251 BP.
XX
AC ABS58082;
XX
DT 05-FEB-2003 (first entry)
XX
DE Agrobacterium tumefaciens btAA gene.
XX
KW Betaine lipid production; betaine lipid compound; DGTS;
KW diacylglycerol-0'-4'-(N,N,N-trimethyl) homoserine; agricultural industry;
KW phosphate-containing fertiliser; transgenic; btAA; gene; ds.
XX
OS Agrobacterium tumefaciens.
XX
FT Key
FT CDS
FT 1. 1251
FT /tag= a
FT /product= "BtA protein"
XX
XX WO200283844-A2.
XX
XX 24-OCT-2002.
XX
XX 09-APR-2002; 2002WO-US011134.
XX
XX 13-APR-2001; 2001US-0283812P.
XX
XX 08-APR-2002; 2002US-00283812.
XX
XX (UNMS) UNIV MICHIGAN STATE.
XX
XX Benning C, Riekhof W, Klug R;
XX
XX WPI; 2003-058632/05.
XX
XX P-PSDB; ABG72178.
XX
XX New composition comprising an isolated and purified DNA molecule, useful
XX for producing Betaine lipids, e.g. diacylglycerol-0'-4'-(N,N,N-trimethyl)
XX homoserine (DGTS) for agricultural applications.
XX
XX Claim 1; Fig 24; 109pp; English.
XX
XX The present invention relates to compositions and methods for producing
XX betaine lipids. The method comprises the expression of recombinant
XX enzymes (e.g. from Rhodospirillum rubrum) in host cells such as
XX bacteria, yeast, and plants to produce betaine lipid compounds including
XX diacylglycerol-0'-4'-(N,N,N-trimethyl) homoserine (DGTS). The methods and
XX compositions of the invention are useful in agricultural applications, the
XX such that the amount of phosphate-containing fertiliser required for the

CC included are an RNA transcribed from the purified DNA, antibodies
 CC produced from the protein, a vector comprising the DNA or nucleic acid, a
 CC host cell comprising the vector, transgenic plants comprising the vector,
 CC and a protein translated from the RNA or encoded by the nucleic acid. The
 CC composition is useful in producing betaine lipid compounds e.g.,
 CC Diacylglycerol-0'-4'-(N,N,N-trimethyl) homoserine (DGTS). The composition
 CC of the invention reduces the amount of phosphate fertilizer needed for
 CC the optimal growth of crop plants. The present sequence encodes a betaine
 CC lipid synthetic enzyme.

XX
 SQ Sequence 1251 BP; 252 A; 423 C; 345 G; 231 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 8,51e-25 Length: 1251
 Score: 326.50 Matches: 115
 Percent Similarity: 47.6% Conservative: 71
 Best Local Similarity: 29.4% Mismatches: 162
 Query Match: 9.4% Indels: 46
 DB: 14 Gaps: 14

US-10-620-914-45 (1-648) x ADW72731 (1-1251)

QY 275 LeuTyrThrGlnSerTrpGluAspProGluProAspMetGluValMetGluLeuLeuAsnPro 294
 DB 130 GTCTATCCGAGATCTGGGAGAGACCCGAGATCGACATGAGAGGATGGAGCTTGGCGAA 189
 QY 295 LysAspThrValLeuThrLeuThrSerGlyGlyCysAsnAlaLeuLeuLeuValGln 314
 DB 190 GGCACCGCATCGTCACCATCGCTCGCGGGCTGCAACATGCTGGCGCTATCTCTCGCGC 249
 QY 315 GlyAlaGlnValValSerValAspCysAsnProAlaGlnSerAlaLeuLeuGluLeu 334
 DB 250 AACCCGGCCAGCATCGATGGTGGACCTCAACCCGACCATCGCGCTGGAACAAGCTG 309
 QY 335 LysLysValAlaLeuGlnLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 353
 DB 310 AAGCTCGCTCGCTCGCCATCTGCCGCCATCAGATGTGTGGCCATCTTGGCCGC 369
 QY 354 GlyValHisProArgLeuGluLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 373
 DB 370 GCCGG-CACCCGAGCAACAGCGTCGTTATGACCGTTTCATCGCGAGCATCTGGATGC 428
 QY 373 nThrSerHisAsnPhetTrpSerLys-----ArgLeuTrpTyrPh 386
 DB 429 CAGCATCAAGGCATCTAGTGTGCAAGCGCACCCCTTTCGGCGCGCTCGCATTTCTGGTGT 488
 QY 386 eGlnHisGlyLeuTyrTrpGlnGlyMetGlyLysLeuCysTrpValLeuGlnCysLe 406
 DB 489 CGACAGACATCTACCGACCGCTGCTCGCGCGCTTTCATCGCGCGCGCCACATCAT 548
 QY 406 uAlaValLeuGlyLeuGlyLysThrValLysArgLeuAlaAsnAlaProThrMetG1 426
 DB 549 GGGCCGCTGCACGGCGTGAATCACC-----GAAATGGCCAAAGACCGGAGCGTGA 602
 QY 426 uGluGlnArgArgLeuTrpAspSerAsnMetLeuLeuHisPheValLysAsnGlyProly 446
 DB 603 CGAACAGCCAGCTTTTTCGACAGCAAGGTGCGCGCTTTTC-----GACAA 650
 QY 446 sProLeuVal---TrpLeuPheValLysPheValSerLeuValLeuPheAsnLysAlaVa 465
 DB 651 GCGGTGTGTGCTGTGCTGACGAGCGCAAGAGCTCGCTT----- 690
 QY 465 lLeuTrpPheGlyGlyValProGlyLysGlnTyr-----AlaLeuLeuLysAl 482
 DB 691 -----TTCCGCTTGGCATTCGCGCGCCAGTATGACGAGCTGGCAAGCTTTCCAG 743
 QY 482 aAspGlyLeuProIleGluAsnTyrIleAlaArgThrMetAspGlyValAlaGluAsnSe 502
 DB 744 CGACGGC---ACGTTTGCCTCCCTCAAGAGCGCTGGAAAGCTTGGCTGCAACTT 800
 QY 502 rHisValArgLysGlnAsnTyrPheTyrTrpAsnCysLeuThrGlyLysPheLeuArgAs 522
 DB 801 CCGGCTC---ACGGACAAATTTATTCGCTTGGCAGGCGCTTGGCGCGCTTATCCCGAGCC 857

QY 522 pAsn-----CysProThrTyrLeuArgGluAlaPheAlaThrLeuLysSerG1 539
 DB 858 GCATGAGGGTCCCTGCGCTTATCTCAAGCCGGAATATTACGAAAGATCCGCAACA 917
 QY 539 yValValAspAsnLeuThrValSerThrAsnPheMetGluGluLeuLys----- 556
 DB 918 CACCGCG---CGCGTCCGCGTGCATCAGCCAGCTATACCGAGCTGTTTCCCGCAAGCC 974
 QY 557 -AlaArgThrTyrThrLysValIleLeuMetAspHisValAspTrpLeuAspMetProVa 576
 DB 975 GGCAAATGGCTCGACCGCTATATCTGCTCGATGCCAGAGCTGGATGACGATGTGCA 1034
 QY 576 lAlaAsnGluLeuAlaGluCysLeuAlaLysGlnValAlaProGlyGlyIleValIleTr 596
 DB 1035 GCTCAACGAGTATGTGCGAGATCAGCCGACCTGCGGCATCGGGGACGCGTCATCTT 1094
 QY 596 pArgSerAlaSerLeu-----SerProTyrAlaGluLeuLeuGlnLys 611
 DB 1095 CCGCACCGCGCGCGA-AAAGAGCGTTATCGAGGCGCGCTTTCGCGCGACATCCGCAACC 1153
 QY 611 sAlaGlyPheAspValArgCysIleArgArgAlaThrGlnGlyTyrMetAspArgValAs 631
 DB 1154 AGTGGTCTATCTCGAAGAGCGCTCCACGA---ACTCAAC---GCCATGACCGCTCGGC 1208
 QY 631 nMetTyrSerSerPheTyrMetAlaArgArg 641
 DB 1209 CATTTATGGCGCTTCCATATCTACCAAGAG 1239

RESULT 15
 ABS58084
 ID ABS58084 standard; DNA; 1251 BP.
 XX
 AC ABS58084;
 DT 05-FEB-2003 (first entry)
 XX
 DE Sinorhizobium meliloti btaA gene.
 KW Betaine lipid production; betaine lipid compound; DGTS;
 KW diacylglycerol-0'-4'-(N,N,N-trimethyl) homoserine; agricultural industry;
 KW phosphate-containing fertiliser; transgenic; btaA; gene; db.
 XX Sinorhizobium meliloti.
 OS
 FH Key Location/Qualifiers
 FT CDS 1..1251
 FT /tag= a
 FT /product= "Bta protein"
 XX WO200283844-A2.
 PN
 XX
 PD 24-OCT-2002.
 XX
 PF 09-APR-2002; 2002WO-US011134.
 XX
 PR 13-APR-2001; 2001US-0283812P.
 PR 08-APR-2002; 2002US-00283812.
 XX
 PA (UNWS) UNIV MICHIGAN STATE.
 PI
 PI Benning C, Riekhof W, Klug R;
 DR WPI: 2003-058632/05.
 DR P-PSDB; ABG72180.
 XX
 PT New composition comprising an isolated and purified DNA molecule, useful
 PT for producing betaine lipids, e.g. Diacylglycerol-0'-4'-(N,N,N-trimethyl)
 PT homoserine (DGTS) for agricultural applications.
 XX
 PS Claim 1; Fig 28; 109pp; English.
 XX The present invention relates to compositions and methods for producing

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 13, 2006, 19:55:21 ; Search time 1196 Seconds
(without alignments)
10849.629 Million cell updates/sec

Title: US-10-620-914-44

Perfect score: 1947

Sequence: 1 atggggctgggtcgtgacgg.....gcgcgaagaagacaactaa 1947

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- N_Geneseq_21.*
- 1: Geneseqn1980s.*
 - 2: Geneseqn1990s.*
 - 3: Geneseqn2000s.*
 - 4: Geneseqn2001as.*
 - 5: Geneseqn2001bs.*
 - 6: Geneseqn2002as.*
 - 7: Geneseqn2002bs.*
 - 8: Geneseqn2003as.*
 - 9: Geneseqn2003bs.*
 - 10: Geneseqn2003cs.*
 - 11: Geneseqn2003ds.*
 - 12: Geneseqn2004as.*
 - 13: Geneseqn2004bs.*
 - 14: Geneseqn2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1947	100.0	1947	14	Adw72747
2	645	33.1	5242	14	Adw72746 Chlamydom
3	100.2	5.1	1252	8	Abes58051 Gene enco
4	100.2	5.1	1252	14	Adw72704 Rhodobact
5	76.8	3.9	2359	12	Adp98536 C. albica
6	73	3.7	2727	14	Adw72752 Neurospor
7	73	3.7	3427	14	Adw72751 Parapoxvi
8	72.4	3.7	137560	12	Abp74816 Sinothizo
9	70	3.6	1251	8	Abes58084
10	70	3.6	1251	14	Adw72735 Sinothizo
11	69.6	3.6	1248	8	Abes58076 Mesorhizo
12	69.6	3.6	1248	14	Adw72725 Mesorhizo
13	69.4	3.6	1092	4	Aah44047 Streptomy
14	69.4	3.6	3003	11	ABD16186 Pseudomon
15	69.4	3.6	3132	11	ABD15936 Pseudomon
16	69.4	3.6	3411	11	ABD16219 Pseudomon
17	67.8	3.5	1092	6	ABL41287 Streptomy
18	67.8	3.5	1104	5	Aah78257 Nucleotid
19	67.8	3.5	1104	6	ABL41308 Streptomy

C	20	67.8	3.5	2000	11	ACL36843	ACL36843 Rice stre
	21	67.8	3.5	6798	4	Aah44043	Aah44043 Streptomy
	22	67.8	3.5	6798	5	Aah78258	Aah78258 Nucleotid
	23	67.8	3.5	6798	6	AAD31022	AAD31022 Streptomy
	24	67.8	3.5	6798	12	ADP90610	ADP90610 Streptomy
	25	67.8	3.5	6798	13	ADR16784	ADR16784 Streptomy
	26	67.8	3.5	8077	6	AAD31028	AAD31028 Operon F
	27	67.8	3.5	8719	12	ADP90614	ADP90614 Streptomy
	28	67.8	3.5	8719	13	ADR16788	ADR16788 Streptomy
	29	67.6	3.5	5858	8	AAD55726	AAD55726 Nephila m
	30	66.6	3.4	1626	12	ADO51196	ADO51196 Staphyloc
	31	66.6	3.4	1626	12	ADO50828	ADO50828 S. coelic
	32	66.6	3.4	1626	13	ADS64696	ADS64696 Streptomy
	33	66.6	3.4	2535	14	ADZ04182	ADZ04182 Modified
	34	65	3.3	11238	10	AAD55817	AAD55817 Micromono
	35	65	3.3	47988	10	ADE86070	ADE86070 Streptomy
	36	65	3.3	60196	10	AAD55810	AAD55810 Micromono
	37	64.8	3.3	2583	14	ADZ07737	ADZ07737 HIV subty
	38	64.6	3.3	2034	14	ADY28052	ADY28052 PERV subc
	39	64.2	3.3	2278	14	ADY57110	ADY57110 Zea mays
	40	64.2	3.3	2807	13	ADK62961	ADK62961 Plant ful
	41	64	3.3	1005	2	AAZ30006	AAZ30006 Optimised
	42	64	3.3	1076	2	AAZ30007	AAZ30007 Optimised
	43	63	3.2	1891	14	ACL70101	ACL70101 M. xanthu
	44	63	3.2	2034	14	ADY28053	ADY28053 PERV subc
	45	63	3.2	2034	14	ACL72051	ACL72051 M. xanthu

ALIGNMENTS

RESULT 1

ADW72747

ID ADW72747 standard; cdna; 1947 BP.

XX ADW72747;

AC ADW72747;

XX 21-APR-2005 (first entry)

XX Chlamydomonas reinhardtii Btal cdna.

XX Phospholipid synthesis; transgenic plant; fertilizer; ss; gene;

XX betaine lipid.

XX Chlamydomonas reinhardtii.

OS Chlamydomonas reinhardtii.

XX Key Location/Qualifiers

FH 1..1947

FT /*tag= a

FT /product= "Btal protein"

XX WO2005009115-A2.

XX 03-FEB-2005.

XX 15-JUL-2004; 2004WO-US022789.

XX 16-JUL-2003; 2003US-00620914.

XX (UNMS) UNIV MICHIGAN STATE.

XX Benning C, Riekhof W;

XX WPI: 2005-112975/12.

XX P-PSDB; ADW72748.

XX New composition comprising a purified DNA having an oligonucleotide

XX sequence encoding a protein, useful in producing Betaine lipid compounds

XX e.g., Diacylglycerol-O-4'-(N,N,N-trimethyl) homoserine (DGTS).

XX Claim 1; SEQ ID NO 44; 147pp; English.

XX The invention relates to a new composition comprising a purified DNA

CC

having an oligonucleotide sequence appearing as ADW72747 (*Chlamydomonas reinhardtii* Btal cDNA) or ADW72752 (*Neurospora crassa* Btal coding region, encoding Btal proteins which are DGTS (diacylglycerol-O-4'-(N,N,N-trimethyl)homoserine) synthetic enzymes. Btal (and BtaA and BtaB) enzymes allow the replacement of phospholipids with non-phosphorus containing lipids in the cell membrane. A transgenic plant expressing the enzymes would have a reduced need for phosphate-containing fertilizer. Also included are an RNA transcribed from the purified DNA, antibodies produced from the protein, a vector comprising the DNA or nucleic acid, a host cell comprising the vector, transgenic plants comprising the vector, and a protein translated from the RNA or encoded by the nucleic acid. The composition is useful in producing betaine lipid compounds e.g., Diacylglycerol-O-4'-(N,N,N-trimethyl) homoserine (DGTS). The composition of the invention reduces the amount of phosphate fertilizer needed for the optimal growth of crop plants. The present sequence encodes a betaine lipid synthetic enzyme.

Sequence 1947 BP: 365 A; 619 C; 613 G; 350 T; 0 U; 0 Other;

Query Match	100.0%	Score 1947;	DB 14;	Length 1947;
Best Local Similarity	100.0%	Pred. No. 0;		
Matches 1947:	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;

[illegible]

QY 1861 CGCGCCACTCAGGCTACATGACCGCGTCAACATGACAGCTCTTCTACATGGCCGC 1920
 DB 1861 CGCGCCACTCAGGCTACATGACCGCGTCAACATGACAGCTCTTCTACATGGCCGC 1920
 QY 1921 CGGAAGGGCGCCAGAGGCAACTAA 1947
 DB 1921 CGGAAGGGCGCCAGAGGCAACTAA 1947

RESULT 2

ADW72746
 ID ADW72746 standard; DNA; 5242 BP.

AC ADW72746;
 XX 21-APR-2005 (first entry)
 XX Chlamydomonas reinhardtii Btal gene.
 XX Phospholipid synthesis; transgenic plant; fertilizer; ds; gene;
 KW betaine lipid.
 XX Chlamydomonas reinhardtii.
 XX WO2005009115-A2.
 XX 03-FEB-2005.
 XX 15-JUL-2004; 2004WO-US022789.
 XX 16-JUL-2003; 2003US-00620914.
 XX (UNMS) UNIV MICHIGAN STATE.
 XX Benning C, Riekhof W;
 XX WPI; 2005-112975/12.
 DR P-PSDB; ADW72748.

XX New composition comprising a purified DNA having an oligonucleotide
 PT sequence encoding a protein, useful in producing Betaine lipid compounds
 PT e.g., Diacylglycerol-0'-4'-(N,N,N,-trimethyl) homoserine (DGTS).

XX Example 4; SEQ ID NO 43; 147pp; English.

XX The invention relates to a new composition comprising a purified DNA
 CC having an oligonucleotide sequence appearing as ADW72747 (Chlamydomonas
 CC reinhardtii Btal cDNA) or ADW72752 (Neurospora crassa Btal coding region,
 CC encoding Btal proteins which are DGTS (diacylglycerol-0'-4'-(N,N,N,-
 CC trimethyl)homoserine) synthetic enzymes. Btal (and BtaA and BtaB) enzymes
 CC allow the replacement of phospholipids with non-phosphorus containing
 CC lipids in the cell membrane. A transgenic plant expressing the enzymes
 CC would have a reduced need for phosphate-containing fertilizer. Also
 CC included are an RNA transcribed from the purified DNA, antibodies
 CC produced from the protein, a vector comprising the DNA or nucleic acid, a
 CC host cell comprising the vector, transgenic plants comprising the vector,
 CC and a protein translated from the RNA or encoded by the nucleic acid. The
 CC composition is useful in producing Betaine lipid compounds e.g.,
 CC Diacylglycerol-0'-4'-(N,N,N,-trimethyl) homoserine (DGTS). The composition
 CC of the invention reduces the amount of phosphate fertilizer needed for
 CC the optimal growth of crop plants. The present sequence encodes a betaine
 CC lipid synthetic enzyme.

XX Sequence 5242 BP; 955 A; 1562 C; 1662 G; 1063 T; 0 U; 0 Other;

Query Match 33.1%; Score 645; DB 14; Length 5242;
 Best Local Similarity 100.0%; Pred. No. 8.1e-110;
 Matches 645; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1045 CAGCTTTCCGCGAGGCGGTGACCCCGCATTTGAGAGCTGTACGAGAGAGCTGGCG 1104
 DB 3228 CAGCTTTCCGCGAGGCGGTGACCCCGCATTTGAGAGCTGTACGAGAGAGCTGGCG 3287

QY 1105 CCCTTCTCTGCGCAACACGACCAACTTCTGTTCCAAAGCGCTCTGGTACTTCTCCAGCAC 1164
 DB 3288 CCCTTCTCTGCGCAACACGACCAACTTCTGTTCCAAAGCGCTCTGGTACTTCTCCAGCAC 3347
 QY 1165 GGCCTGTACTACCAAGGCGGCATGGGCAAGCTGTGCTGGGTGCTGAGTGGCTGGCGGTG 1224
 DB 3348 GGCCTGTACTACCAAGGCGGCATGGGCAAGCTGTGCTGGGTGCTGAGTGGCTGGCGGTG 3407
 QY 1225 GTGCTGGGACTGGGCAAGACCGCTCAAGCGCTCGCCAAACGCGCCCAAAATGGAGGAGCAG 1284
 DB 3408 GTGCTGGGACTGGGCAAGACCGCTCAAGCGCTCGCCAAACGCGCCCAAAATGGAGGAGCAG 3467
 QY 1285 CGCCGTCTGTGGGACAGCAACATGCTCATCTTCTGTAAGAACCGGCGCCAAAGCCGCTG 1344
 DB 3468 CGCCGTCTGTGGGACAGCAACATGCTCATCTTCTGTAAGAACCGGCGCCAAAGCCGCTG 3527
 QY 1345 GTGCTGGCTGTGCTCAAGTTCTGAGGCTGTGCTCTTCAACAAAGCGCGTGTGTGTTTC 1404
 DB 3528 GTGCTGGCTGTGCTCAAGTTCTGAGGCTGTGCTCTTCAACAAAGCGCGTGTGTGTTTC 3587
 QY 1405 GCGGCGGCGGTGCGGCGCAAGCAGTACGCGCTGATCAAGCGGACGCGCATCCCATTTGAG 1464
 DB 3588 GCGGCGGCGGTGCGGCGCAAGCAGTACGCGCTGATCAAGCGGACGCGCATCCCATTTGAG 3647
 QY 1465 AACTATATCGCGCGCACCATGGAACGCGCTGGCGGAGAACTCGCACCTGCGCAAGCAGAAC 1524
 DB 3648 AACTATATCGCGCGCACCATGGAACGCGCTGGCGGAGAACTCGCACCTGCGCAAGCAGAAC 3707
 QY 1525 TACTTCTACTACACTGCTCCCGCAAGTTCTTCTGGGAGCAACTGCGCCACCTACCTG 1584
 DB 3708 TACTTCTACTACACTGCTCCCGCAAGTTCTTCTGGGAGCAACTGCGCCACCTACCTG 3767
 QY 1585 GCGGAGGCGGCTTTCGCCACCTCAAGAGTGGGTGGTGGACAACTGACCGCTCTCCACC 1644
 DB 3768 GCGGAGGCGGCTTTCGCCACCTCAAGAGTGGGTGGTGGACAACTGACCGCTCTCCACC 3827
 QY 1645 AACTTCTTTCATGAGGAGCTCAAAGCGCGCACCTTACACCAAGGTG 1689
 DB 3828 AACTTCTTTCATGAGGAGCTCAAAGCGCGCACCTTACACCAAGGTG 3872

RESULT 3

ABSS8051
 ID ABSS8051 standard; DNA; 1252 BP.

XX AC ABSS8051;
 XX 05-FEB-2003 (first entry)
 XX Gene encoding Rhodobacter sphaeroides btaA protein.
 KW Betaine lipid production; betaine lipid compound; DGTS;
 KW diacylglycerol-0'-4'-(N,N,N,-trimethyl) homoserine; agricultural industry;
 KW phosphate-containing fertilizer; transgenic; btaA; gene; ds.
 XX Rhodobacter sphaeroides.
 XX Key Location/Qualifiers
 FH 1..1251
 CDS /tag= a
 FT /partial
 FT /product= "BtaA protein"
 FT /transl_except= (pos:1..3, aa:Met)
 FT /note= "This sequence lacks a start codon"

XX WO200283844-A2.

XX 24-OCT-2002.

XX 09-APR-2002; 2002WO-US011134.

XX 13-APR-2001; 2001US-0283812P.

```
PR 08-APR-2002; 2002US-00283812.
XX
PA (UNMS ) UNIV MICHIGAN STATE.
XX
PI Benning C, Riekhof W, Klug R;
XX
DR WPI; 2003-058632/05.
DR P-PSDB; ABG72176.
XX
XX New composition comprising an isolated and purified DNA molecule, useful
PT for producing Betaine lipids, e.g. Diacylglycerol-0-4'-(N,N,N-trimethyl)
PT homoserine (DGTS) for agricultural applications.
XX
PS Claim 1; Fig 7; 109pp; English.
XX
CC The present invention relates to compositions and methods for producing
CC betaine lipids. The method comprises the expression of recombinant
CC enzymes (e.g. from Rhodobacter sphaeroides) in host cells such as
CC bacteria, yeast, and plants to produce betaine lipid compounds including
CC diacylglycerol-0-4'-(N,N,N-trimethyl) homoserine (DGTS). The methods and
CC compositions of the invention are useful in agricultural applications,
CC such that the amount of phosphate-containing fertilizer required for the
CC growth of a particular plant is decreased. The polynucleotide sequences
CC encoding the recombinant enzymes may be used to produce vectors which can
CC be used to produce transgenic plants. The present sequence encodes R.
CC sphaeroides btaA protein
XX
SQ Sequence 1252 BP; 152 A; 488 C; 424 G; 188 T; 0 U; 0 Other;

Query Match 5.1%; Score 100.2; DB 8; Length 1252;
Best Local Similarity 50.9%; Pred. No. 3.8e-09;
Matches 299; Conservative 0; Mismatches 273; Indels 15; Gaps 2;

Qy 809 TCCGCGCCACCTTCTCTACACGAGTCGTGGAGGACCCCGAGCCGGATATGAGGTGA 868
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 869 TGGAGATCAACCCCAAGACACGCTGCTGACCTGACTAGCGGGCTGCAATGCCCTGA 928
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 182 TCGGCATCCGCGCCCGGGACCGCTGTGTGGCAATCGCTCGGGCGGTTCACGTGCTT 241
Qy 929 ACCTGTGTGTGACAGGGCGCGCCAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 988
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 242 CCTATCTCACGAGGGCGCGGCTCGATCTCGCGGTGGATCTTCGCGCGCCCATGTGG 301
Qy 989 CGCTTCTGAGCTGAAGAGGTGGCCATTGACGAGCTG---GAGTTTGAAGACGTGGC 1045
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 302 CGCTGGGGCGGCTGAAGCTCGCCCGCGCGGAGCGCTGCCCGACCATGCCGCTTCTTCG 361
Qy 1046 AGCTGTTCCGCGAGGGCGGTGCACCCGCGCATTTGAGGAGCTGTACGAGAAGCTGGCG 1105
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 362 ATCTCTTCGCTGCGCGAGACCTGCGCGGCAATGCGGCGCTCTACGACGGCACATCGCG 421
Qy 1106 CTTTCTCTGTGCACAAACAGCCACAACTTCTGGTTCACAGCGCTCTGGTACTTCCAGACG 1165
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 422 CCGCGCTCGAGCGCGGAGCGCGCTACTTGGAGGCGCGAGCCCTTCGGCGGCGCA 481
Qy 1166 GCCTGTACTACAGGGCGGCAATGGGCAAGCTGTCTGGGTGCTGCACTGCTGGCGG--- 1222
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 482 TCCAGCTGTTTCGAGCGGGCTTCTTACCGGACGCGTTCGCGCGCTTTCATCGCGCGG 541
Qy 1223 -----TGGTCTCGGACTGGGCAAGACCGTCAAGCGCTCGCCCAAGCGCCCAAA 1273
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 542 CCCATAGCTCGCGGGCGCGGGCACCGACCTGCGGGGCTTTCTCGACTGTCCCGACA 601
Qy 1274 TGGAGGAGCAGCGCGCTGTGTGGACAGCAACATGCTCATCCACTTCTGTGAAGAACGGCG 1333
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 602 TCGAGGCGCAGCGCAGCTTCTTCTACGCCCATATCGGGCGCTCTTCGAGGCGCGGTGG 661
Qy 1334 CCAAGCGCTGTGGTGTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1380
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 662 TGCAGGCGCTCGCCGACGCGCGCGCGCTCTTTCGGGCTGGGATC 708
```

```
RESULT 4
ADW72704
XX
XX ADW72704 standard; cDNA; 1252 BP.
AC
XX ADW72704;
DT
XX 21-APR-2005 (first entry)
XX
XX Rhodobacter sphaeroides btaA gene, cDNA.
DE
XX Phospholipid synthesis; transgenic plant; fertilizer; ss; gene;
KW betaine lipid.
XX
XX Rhodobacter sphaeroides.
XX
XX Key Location/Qualifiers
CDS 1..1251
FT /*tag= a
FT /product= "BtaA protein"
FT /trans_except= (pos:1..3,aa:Met)
XX
XX W02005009115-A2.
XX
XX 03-FEB-2005.
XX
XX 15-JUL-2004; 2004WO-US022789.
XX
XX 16-JUL-2003; 2003US-00620914.
XX
XX (UNMS ) UNIV MICHIGAN STATE.
XX
XX Benning C, Riekhof W;
XX
XX WPI; 2005-112975/12.
DR P-PSDB; ADW72706.
XX
XX New composition comprising a purified DNA having an oligonucleotide
PT sequence encoding a protein, useful in producing Betaine lipid compounds
PT e.g., Diacylglycerol-0-4'-(N,N,N-trimethyl) homoserine (DGTS).
XX
XX Example 3; SEQ ID NO 1; 147pp; English.
XX
XX The invention relates to a new composition comprising a purified DNA
XX having an oligonucleotide sequence appearing as ADW72747 (Chlamydomonas
XX reinhardtii Btal cDNA) or ADW72752 (Neurospora crassa Btal coding region,
XX encoding Btal proteins which are DGTS (diacylglycerol-0-4'-(N,N,N-
XX trimethyl)homoserine) synthetic enzymes. Btal (and BtaA and BtaB) enzymes
XX allow the replacement of phospholipids with non-phosphorus containing
XX lipids in the cell membrane. A transgenic plant expressing the enzymes
XX would have a reduced need for phosphate-containing fertilizer. Also
XX included are an RNA transcribed from the purified DNA, antibodies
XX produced from the protein, a vector comprising the DNA or nucleic acid, a
XX host cell comprising the vector, transgenic plants comprising the vector,
XX and a protein translated from the RNA or encoded by the nucleic acid. The
XX composition is useful in producing Betaine lipid compounds e.g.,
XX Diacylglycerol-0-4'-(N,N,N-trimethyl) homoserine (DGTS). The composition
XX of the invention reduces the amount of phosphate fertilizer needed for
XX the optimal growth of crop plants. The present sequence encodes a betaine
XX lipid synthetic enzyme.
SQ Sequence 1252 BP; 152 A; 488 C; 424 G; 188 T; 0 U; 0 Other;

Query Match 5.1%; Score 100.2; DB 14; Length 1252;
Best Local Similarity 50.9%; Pred. No. 3.8e-09;
Matches 299; Conservative 0; Mismatches 273; Indels 15; Gaps 2;

Qy 809 TCCGCGCCACCTTCTCTACACGAGTCGTGGAGGACCCCGAGCCGGATATGAGGTGA 868
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 122 TCTTCCACGGCTCGTCTATCCGAGATCTGGAGGATCCGGCGGTGGACATGCGCGCC 181
Qy 869 TGGAGATCAACCCCAAGACACGAGTCGTGTGCTGACCTGTAGCGGGCTGCAATGCCCTGA 928
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Db 182 TCGCATCCGCGCGGACCGGCTGTGTGCGCATCGCTCGGGGGTTCACAGTGTCTT 241
Qy 929 ACTGTGCTGTGACGGGGCGCGGACGAGTGTGTGTGTGAGTGTCAACCCGCGCAGTCGG 988
Db 242 CCTATCTCACGACGGGCGCGGCTCGATCTCGCGGTGGATCTCTCGCGCGCCCATGTGG 301
Qy 989 CGCTTCTGAGAGCTGAAGAGGTGGCCATTACAGAGCTG---GAGTTGAGAGAGTGTGGC 1045
Db 302 CGCTGGGGCGGCTGAAGCTCGCGCGCGCGGACGCTGCCCGACCATGCGCGCTTCTTCG 361
Qy 1046 AGCTGTTTCGCGAGGGCGGTGCACCGCGCATTTGAGGAGCTGTACGAGAAGCTGCGCG 1105
Db 362 ATCTTCTGTGCGCGACAGCTGCCGCGCAATGCGGCGCTTACGACCGCCATCGCGC 421
Qy 1106 CTTTCTGTGCGCAACAGCCACAACTTCTGTGTCGAAGCGCTCTGTGTACTTCCAGCAG 1165
Db 422 CCGCGCTCGACGGCGCGGCGCGCTTCTACTGTGGAGGCGCGCAGCCCTTCGCGCGCGCA 481
Qy 1166 GCCTGTACTACGAGGGCGCATGGGCAAGCTGTGCTGGGTGTGCTGAGTGTCTGCGCG 1222
Db 482 TCCAGCTGTTCGAGCGCGGCTTCTACCGGACCGTGTGCCCTTCGCGCGCTTTCATCG 541
Qy 1223 -----TGTGTCTGGGACTGGGCAAGACCGTCAAGCGCTCGCCCAAGCGCCACAA 1273
Db 542 CCATAGCTCGCGCGCGCGCGGCGGCGGCGGCGGCGGCTTCTGAGTGTCCGACGA 601
Qy 1274 TGAGGAGCAGCGCGCTCTGTGGGACAGCAACATGTCTCATCCACTTGTGGAAGACGGCG 1333
Db 602 TCGAGGCGCAGCGCAGCTTCTTACGCCCATATCGGGCGCGCTTCGAGGCGCGCGTGG 661
Qy 1334 CCAAGCGCGTGTGTGCTGTTCTGTCAGTTCGTGAGCTGTGTGCTC 1380
Db 662 TGCAGGCGCTCGCGCGCGCGCGCGCGCGCTTCTTCGGGCTGGGGATC 708

RESULT 5

ADP98536
ID ADP98536 standard; DNA; 2259 BP.

XX AC ADP98536;

XX DT 23-SEP-2004 (first entry)

XX DE C. albicans specific gene, orf6.3438, DNA sequence.

XX KW Diploid fungal cell; allele; gene disruption cassette;
XX KW promoter replacement fragment; antifungal; fungicide; gene therapy;
XX KW infection; Candida albicans; gene; ds.

XX OS Candida albicans.

XX PN WO2004056965-A2.

XX PD 08-JUL-2004.

XX PF 19-DEC-2003; 2003WO-US040618.

XX PR 19-DEC-2002; 2002US-0434832P.

XX XX (ELIT-) ELITRA PHARM INC.

XX PA (ELIT-) ELITRA CANADA LTD.

XX PI Roemer T, Jiang B, Boone C, Bussey H;

XX DR WPI: 2004-500296/47.

XX DR P-PSDB; ADP98846.

XX PT Constructing a strain of diploid fungal cells in which both alleles of a
XX PT gene are modified comprises modifying the alleles of a gene in the fungal
XX PT cells by recombination using a gene disruption cassette and a promoter
XX PT replacement fragment.

XX PS Claim 33; SEQ ID NO 6021; 163pp; English.

XX

The invention relates to a novel method for constructing a strain of diploid fungal cells in which both alleles of a gene are modified. The method comprises modifying the alleles of a gene in diploid fungal cells by recombination using a gene disruption cassette and a promoter replacement fragment. The invention further comprises: assembling a collection of diploid fungal cells each of which comprises modified alleles of a different gene; a strain of diploid fungal cells comprising modified alleles of a gene, where the first allele of the gene is inactivated by a gene disruption cassette comprising a nucleotide sequence encoding an expressible selectable marker; and the expression of the second allele of the gene is regulated by a heterologous promoter that is operably linked to the coding region of the second allele of the gene, and where the gene encodes the polypeptide mentioned above; a collection of diploid fungal strains comprising the diploid strains cited above, where substantially all the different genes that encode the above amino acid sequences are modified and are present in different diploid strains in the collection; a nucleic acid molecule microarray comprising nucleic acid molecules, where each nucleic acid molecule comprises a nucleotide sequence that is hybridizable to a target nucleotide sequence comprising any of the 310 nucleotide sequences listed in the specification (ADP98516-ADP98825); identifying a gene that is essential to the survival or growth of a fungus, that contributes to the virulence and/or pathogenicity of a fungus, or that contributes to the resistance of a diploid fungus to an antifungal agent; identifying an antifungal agent that inhibits the growth of a diploid fungus, or a therapeutic levels of proteins or gene transcripts with the inhibition of growth or proliferation of a diploid fungal cell; a purified or isolated nucleic acid molecule comprising a nucleotide sequence encoding a gene product required for proliferation of Candida albicans, where the gene product consists of any of the above-mentioned amino acid sequences; a vector comprising a promoter operably linked to the nucleic acid molecule cited above; a host cell containing the vector; a purified or isolated polypeptide comprising any of the 61 amino acid sequences given in the specification (ADP96718-ADP96779); a fusion protein comprising a fragment of a first polypeptide fused to a second polypeptide, the fragment consisting of at least 6 consecutive residues of any of ADP98826-ADP99135; producing a polypeptide; identifying a compound which modulates the activity of a gene product encoded by a nucleic acid comprising any of ADP98516-ADP98825; eliciting an immune response in an animal; a strain of Candida albicans, where a first allele of a gene comprising any of ADP98516-ADP98825 is inactive and a second allele of the gene is under the control of a heterologous promoter; identifying a compound or binding partner that binds to the polypeptide comprising any of ADP98826-ADP99135, or its fragment; identifying a compound having the ability to inhibit growth or proliferation of Candida albicans; inhibiting growth or proliferation of Candida albicans cells; manufacturing an antimycotic compound; treating an infection of a subject by Candida albicans; preventing or containing contamination of an object by Candida albicans, or for preventing or inhibiting formation on a surface of a biofilm comprising Candida albicans; a pharmaceutical composition comprising a therapeutic amount of an agent which reduces the activity or level of a gene product encoded by a nucleic acid comprising any of ADP98516-ADP98825 in a pharmaceutical carrier; an antibody preparation which binds the polypeptide; methods for evaluating a compound against a target gene product encoded by any of ADP98516-ADP98825; identifying an antimycotic compound; a computer or a computer readable medium that comprises at least one of the nucleotide sequences mentioned in the specification or at least one amino acid sequence selected from ADP98826-ADP99135; a method assisted by a computer for identifying a putatively essential gene of a fungus; and a protein array comprising proteins, where at least one protein comprises an amino acid sequence or a portion of an amino acid sequence selected from ADP98516-ADP98825. The novel methods and compositions have fungicide activity. The compositions may be used in gene therapy. The composition and methods are useful for drug screening purposes or for diagnosing, preventing or treating infections associated with Candida albicans. These may also be used for constructing strains useful for identification and validation of gene products as effective targets for therapeutic intervention, for identifying and validating gene products as effective targets for therapeutic intervention, and for collecting identified essential genes. This polynucleotide sequence represents a Candida albicans fungal specific gene of the invention.

CC NOTE: This sequence was downloaded from an electronic sequence listing
CC provided on the WIPO website.

XX	Sequence	2259 BP;	714 A;	359 C;	406 G;	780 T;	0 U;	0 Other;	
SQ	Query Match	3.9%;	Score	76.8;	DB	12;	Length	2259;	
	Best Local Similarity	54.3%;	Pred. No.	8.5e-05;					
	Matches	178;	Conservative	0;	Mismatches	147;	Indels	3;	Gaps 1;
Qy	212	ACCTCATCTGGGTGGACCTGGTGGTGGCGCACTGGGGAGAAATGTCGATATGATGGCTGATT	271						
Db	323	ATCTTATTGGATTGATATTGGTGGTGGAACTGGTCCATATTTGAATTCATGGATGAAA	382						
Qy	272	ACATCGACCT---GGCGAAGTTCAAGTCCATCTACGTGGTGCAGCTGTGCCACTGCCTGCT	328						
Db	383	TTAGTAAATATCTGAAAACTTTAAAGCTGTTTATTGGTTGATCTTTCCCACTCTTTGT	442						
Qy	329	GCGAGGTGGCCAGAAAGACGCGAAGGCTGGAAGGCTGGAAGAATGTCCAGGTCTGTGGAGG	388						
Db	443	GTGAAGTTGCTTAAGGCAAGATTTGAAGGCCCATGAATGACAAATGTTTCATGTATTAGTTG	502						
Qy	389	CCGACGCTGCGCAATTTTCGGCCCCCTGAGGGCACGGGACCGCTCATACCTTCTTCCTACT	448						
Db	503	CTGATGCTGTGATTTTACTATTGATTATGATAGTGTGATTTCGATTACTTTTCTTTATT	562						
Qy	449	CGCTCAGATGATTTCCACCGTTTCCAAACGTCATCGACCAGGCTTGCTCGTACCTGCCCC	508						
Db	563	CATTGTCGATGATCCCAACTTTCATGTGCTATCGATAATGCTGTTTCTAAATTAGATA	622						
Qy	509	AAGACGGCCTGGTGGGCTTGCCGACTT	536						
Db	623	TGGAAGGTATTATTGGCACTGTGGATT	650						

RESULT 6	
ADW72752	
ID	ADW72752 standard; DNA; 2727 BP.
XX	
AC	ADW72752;
XX	
DT	21-APR-2005 (first entry)
XX	
DE	Neurospora crassa Btal coding region.
XX	
KW	Phospholipid synthesis; transgenic plant; fertilizer; ss; gene;
KW	betaine lipid.
XX	
OS	Neurospora crassa.
XX	
Key	Location/Qualifiers
FT	1. .2727
CDS	/*tag= a
FT	/product= "Btal"
FT	
XX	
XX	WO2005009115-A2.
PN	
XX	
XX	03-FEB-2005.
PD	
XX	
PF	15-JUL-2004; 2004WO-US022789.
XX	
XX	16-JUL-2003; 2003US-00620914.
XX	
XX	(UNMS) UNIV MICHIGAN STATE.
PA	
XX	
PI	Benning C, Riekhof W;
XX	
PI	WPI; 2005-112975/12.
DR	P-PSDB; ADW72753.
DR	
XX	
XX	New composition comprising a purified DNA having an oligonucleotide
FT	sequence encoding a protein, useful in producing Betaine lipid compounds
FT	e.g., Diacylglyceryl-O-4'-(N,N,N,-trimethyl) homoserine (DGTS).
PT	
XX	

Claim 1; SEQ ID NO 49; 147pp; English.

The invention relates to a new composition comprising a purified DNA having an oligonucleotide sequence appearing as ADW72747 (Chlamydomonas reinhardtii Btal cDNA) or ADW72752 (Neurospora crassa Btal coding region, encoding Btal proteins which are DGTS (diacylglycerol-0'-4'-(N,N,N-trimethyl)homoserine) synthetic enzymes. Btal (and BtaA and BtaB) enzymes allow the replacement of phospholipids with non-phosphorus containing lipids in the cell membrane. A transgenic plant expressing the enzymes would have a reduced need for phosphate-containing fertilizer. Also included are an RNA transcribed from the purified DNA, antibodies produced from the protein, a vector comprising the DNA or nucleic acid, a host cell comprising the vector, transgenic plants comprising the vector, and a protein translated from the RNA or encoded by the nucleic acid. The composition is useful in producing Betaine lipid compounds e.g., Diacylglycerol-0'-4'-(N,N,N-trimethyl) homoserine (DGTS). The composition of the invention reduces the amount of phosphate fertilizer needed for the optimal growth of crop plants. The present sequence encodes a betaine lipid synthetic enzyme.

Sequence 2727 BP; 616 A; 766 C; 713 G; 632 T; 0 U; 0 Other;

Query Match 3.7%; Score 73; DB 14; Length 2727;
 Best Local Similarity 51.0%; Pred. No. 0.00044;
 Matches 172; Conservative 0; Mismatches 165; Indels 0; Gaps 0;

QY 779 ACGAGGAGCGGTGGAGCGGCCCGCCCATGTTCCCGCCACCTTCCTGTATCACGCAATCGT 838
 DB 1448 ACGACGATCAACTCCCGAAGCACACCCAGTTCAATGACGAGTACATCTACGCTTTACCT 1507

QY 839 GGGAGGACCCGAGCCGGATATGGAGGTGATGAGATCAACCCCAAGGACACGGTGCTGA 898
 DB 1508 GCGAAGACTCGCGCTCGACAGAGAACTCCTTAACCTCGGGCCCGACGACGTGCTCCTAG 1567

QY 899 CCTGTACTAGCGCGCGCTGCAATGCCCTGAACCTGTGGTCAGGGGGCGGCACGCTGG 958
 DB 1568 CCATCAACGACGCGCGCGCAACATCTTTTCCTGATGACAGAGTCCCGCTGCGGTGC 1627

QY 959 TGTTCGGTGGACTGCAACCCCGCCAGTCGCGCGTTCCTGGAGCTGAAGAAGTGGCCATTC 1018
 DB 1628 ACGGCATCGACTTAACCCAGCCCAAAACCACTGCTTGAACCTCAAAGTCGCTCTTTTA 1687

QY 1019 AGCAGCTGGAGTTTGAGGACGTGTGGCAGCTGTTCGGCGAGGGCGTGCACCCCGGCATTG 1078
 DB 1688 CGACTCTGGATTACCCCGACGTCTGGAAGATCTTCGGTGGAGGCGAAACACCCCGACTTTC 1747

QY 1079 AGGAGCTGTACGAGAGAGAGCTGCGGCCCTTCCTCTGTC 1115
 DB 1748 GTCACGTGCTCATCTCCAAACTCTCCCTCACCTCTC 1784

RESULT 7
 ADW72751
 ID ADW72751 standard; DNA; 3427 BP.
 XX AC
 XX ADW72751;
 XX AC
 XX 21-APR-2005 (first entry)
 XX AC
 XX Neurospora crassa Btal gene.
 XX AC
 XX Phospholipid synthesis; transgenic plant; fertilizer; ds; gene;
 XX AC
 XX betaine lipid.
 XX AC
 XX Neurospora crassa.
 XX AC
 XX WO2005009115-A2.
 XX AC
 XX 03-FEB-2005.
 XX AC
 XX 15-JUL-2004; 2004WO-US022789.
 XX AC
 XX 16-JUL-2003; 2003US-00620914.
 XX AC


```
FT      CDS      encoded protein is not provided"
FT      complement(8070..9989)
FT      /*tag= l
FT      /product= "Parapoxvirus ovis polypeptide"
FT      /note= "This ORF is specifically claimed within claim 3
FT      of the specification, although the sequence of the
FT      encoded protein is not provided"
FT      complement(10062..11195)
FT      /*tag= m
FT      /product= "Parapoxvirus ovis polypeptide"
FT      /note= "This ORF is specifically claimed within claim 3
FT      of the specification, although the sequence of the
FT      encoded protein is not provided"
FT      11227..11493
FT      /*tag= n
FT      /product= "Parapoxvirus ovis polypeptide"
FT      /note= "This ORF is specifically claimed within claim 3
FT      of the specification, although the sequence of the
FT      encoded protein is not provided"
FT      11802..12038
FT      /*tag= o
FT      /product= "Parapoxvirus ovis polypeptide"
FT      /note= "This ORF is specifically claimed within claim 3
FT      of the specification, although the sequence of the
FT      encoded protein is not provided"
FT      complement(12080..12358)
FT      /*tag= p
FT      /product= "Parapoxvirus ovis polypeptide"
FT      /note= "This ORF is specifically claimed within claim 3
FT      of the specification, although the sequence of the
FT      encoded protein is not provided"
FT      complement(12364..13980)
FT      /*tag= q
FT      /product= "Parapoxvirus ovis polypeptide"
FT      /note= "This ORF is specifically claimed within claim 3
FT      of the specification, although the sequence of the
FT      encoded protein is not provided"
FT      complement(14053..14826)
FT      /*tag= r
FT      /product= "Parapoxvirus ovis polypeptide"
FT      /note= "This ORF is specifically claimed within claim 3
FT      of the specification, although the sequence of the
FT      encoded protein is not provided"
FT      15080..15394
FT      /*tag= s
FT      /product= "Parapoxvirus ovis polypeptide"
FT      /note= "This ORF is specifically claimed within claim 3
FT      of the specification, although the sequence of the
FT      encoded protein is not provided"
FT      complement(15423..16838)
FT      /*tag= t
FT      /product= "Parapoxvirus ovis polypeptide"
FT      /note= "This ORF is specifically claimed within claim 3
FT      of the specification, although the sequence of the
FT      encoded protein is not provided"
FT      complement(16847..19021)
FT      /*tag= u
FT      /product= "Parapoxvirus ovis polypeptide"
FT      /note= "This ORF is specifically claimed within claim 3
FT      of the specification, although the sequence of the
FT      encoded protein is not provided"
FT      19156..19704
FT      /*tag= v
FT      /product= "Parapoxvirus ovis polypeptide"
FT      /note= "This ORF is specifically claimed within claim 3
FT      of the specification, although the sequence of the
FT      encoded protein is not provided"
FT      complement(19736..20314)
FT      /*tag= w
FT      /product= "Parapoxvirus ovis polypeptide"
FT      /note= "This ORF is specifically claimed within claim 3
FT      of the specification, although the sequence of the
FT      encoded protein is not provided"

CDS      20401..22101
/*tag= x
/product= "Parapoxvirus ovis polypeptide"
/note= "This ORF is specifically claimed within claim 3
of the specification, although the sequence of the
encoded protein is not provided"
22125..22940
/*tag= y
/product= "Parapoxvirus ovis polypeptide"
/note= "This ORF is specifically claimed within claim 3
of the specification, although the sequence of the
encoded protein is not provided"
23003..23866
/*tag= z
/product= "Parapoxvirus ovis polypeptide"
/note= "This ORF is specifically claimed within claim 3
of the specification, although the sequence of the
encoded protein is not provided"
complement(23873..26908)
/*tag= aa
/product= "Parapoxvirus ovis polypeptide"
/note= "This ORF is specifically claimed within claim 3
of the specification, although the sequence of the
encoded protein is not provided"
26926..27213
/*tag= ab
/product= "Parapoxvirus ovis polypeptide"
/note= "This ORF is specifically claimed within claim 3
of the specification, although the sequence of the
encoded protein is not provided"
complement(27216..27626)
/*tag= ac
/product= "Parapoxvirus ovis polypeptide"
/note= "This ORF is specifically claimed within claim 3
of the specification, although the sequence of the
encoded protein is not provided"
complement(27616..29754)
/*tag= ad
/product= "Parapoxvirus ovis polypeptide"
/note= "This ORF is specifically claimed within claim 3
of the specification, although the sequence of the
encoded protein is not provided"
complement(29800..32217)
/*tag= ae
/product= "Parapoxvirus ovis polypeptide"
/note= "This ORF is specifically claimed within claim 3

Query Match      3.7%; Score 72.4; DB 12; Length 137560;
Best Local Similarity 42.9%; Pred. No. 0.00081;
Matches 461; Conservative 0; Mismatches 611; Indels 2; Gaps 2;

QY      618      CGACAACTTGACATCGGCCCGGCGCGCTTACCTCGAGCAGAGCTGGAGCGCGT 677
Db      20613    CGAGGCTTTCGAGCGCGTGTCTCAAGGCGGCTTCGAGCGCCCTAGCTCTTTGTGCGGC 20672
QY      678      GTGGGAGCAGAACACCCAGGGTTCGATCCCTACGTCCGCTGCTGCGCGCCCTACTA 737
Db      20673    GCTCTCGGCGCATTTGAGGCCTTGTGCGGCATCCGGGCTTACCGCGGACCTCGCG 20732
QY      738      CGTGTGATTTGGCGGCTGCCAGCGTTGGCCACGCCCTCGACGAGAGCGCGTGGAGCG 797
Db      20733    CGAGCTGTCTGCCGACGCGCGCTTCATGAGAGTGGTTCGCGCGCGCCGAGCTCGCGGA 20792
QY      798      GCGGCCCATTTCCCGCCACCTTCTGTACAGCAGCTGTGGGAGAGCCCGAGCGCGA 857
Db      20793    GGTCAATTTGGCGTGAACACGACATTCGCCGCGAACCCGCTCTTCGCGAGCGGAGCCCGT 20852
QY      858      TATGGAGGTGATGAGATCAACCCCAAGGACACGGTGTCTGACCCCTGACTAGCGCGGCTG 917
Db      20853    GCGGACGCGGAGTCAATTTTCGCAAGACCTTCGCAAGACCGAGTTCGCGCGCTCAA 20912
QY      918      CAATGCCCTGAACCTGTGTGTGAGGGGCGCGCGAGTGGTGTCTGCTGACTCAACCC 977
```


OS Sinorhizobium meliloti.
 XX Key Location/Qualifiers
 FH CDS 1. .1251
 FT /*tag= a
 FT /product= "BtaA protein"
 XX
 XX WO200509115-A2.
 XX
 XX 03-FEB-2005.
 XX
 XX 15-JUL-2004; 2004WO-US022789.
 XX
 XX 16-JUL-2003; 2003US-00620914.
 XX
 XX (UNMS) UNIV MICHIGAN STATE.
 XX
 XX Benning C, Riekhof W;
 XX
 XX WPI; 2005-112975/12.
 DR P-PSDB; ADW72736.
 DR
 XX
 XX New composition comprising a purified DNA having an oligonucleotide
 PT sequence encoding a protein, useful in producing Betaine lipid compounds
 PT e.g., Diacylglycerol-0-4'-(N,N,N-trimethyl) homoserine (DGTS).
 XX
 XX Disclosure; SEQ ID NO 32; 147bp; English.
 PS
 XX
 XX The invention relates to a new composition comprising a purified DNA
 CC having an oligonucleotide sequence appearing as ADW72747 (Chlamydomonas
 CC reinhardtii Btal cDNA) or ADW72752 (Neurospora crassa Btal coding region,
 CC encoding Btal proteins which are DGTS (diacylglycerol-0-4'-(N,N,N-trimethyl)homoserine) synthetic enzymes. Btal (and BtaA and BtaB) enzymes
 CC allow the replacement of phospholipids with non-phosphorus containing
 CC lipids in the cell membrane. A transgenic plant expressing the enzymes
 CC would have a reduced need for phosphate-containing fertilizer. Also
 CC included are an RNA transcribed from the purified DNA, antibodies
 CC produced from the protein, a vector comprising the DNA or nucleic acid, a
 CC host cell comprising the vector, transgenic plants comprising the vector,
 CC and a protein translated from the RNA or encoded by the nucleic acid. The
 CC composition is useful in producing Betaine lipid compounds e.g.,
 CC Diacylglycerol-0-4'-(N,N,N-trimethyl) homoserine (DGTS). The composition
 CC of the invention reduces the amount of phosphate fertilizer needed for
 CC the optimal growth of crop plants. The present sequence encodes a betaine
 CC lipid synthetic enzyme.
 XX
 XX Sequence 1251 BP; 227 A; 440 C; 355 G; 229 T; 0 U; 0 Other;
 SQ
 Query Match 3.6%; Score 70; DB 14; Length 1251;
 Best Local Similarity 52.8%; Pred. No. 0.0015;
 Matches 151; Conservative 0; Mismatches 135; Indels 0; Gaps 0;
 QY 806 TGTTCGCCGCCACCTTCTGTACACGAGTGTGTGGAGGACCCCGAGCGGATATGAGG 865
 DB 116 TGCTCTTTCCGACTCGTCTACCCGAGATCTCGGAGGACCCGATTTGTGACATGGAAG 175
 QY 866 TGATGGAGATCAACCCCAAGGACAGGTGTGACCCCTGACTAGCGGGCTGCAATGCC 925
 DB 176 CGATGCGATCCGTCCTCCGGACATCGGATCGTGACGATCGGTTCCGGCGGCTGCAATGC 235
 QY 926 TGAACCTGCTGTGCAGGGGCGGCGAGGTGTGTGCTGGTGAATGCAACCCCGCGAGT 985
 DB 236 TGACCTATCTCTCCGCGAGGCTGCCGAGATAGACGTGTGTGATCTCAACCCCATCACA 295
 QY 986 CGGCGCTTCTGGAGCTGAAGAAGTGGCCATTACAGAGCTGGAGTTTGGAGACGTGTGGC 1045
 DB 296 TCGCGCTCAACCGGCTGAAGCTGTCTGCTTTCCGACCTGCGGAGCCACAAGACGTGG 355
 QY 1046 AGCTGTTCCGGCGAGGGGTGACCCCGGCAATTGAGGAGCTGTACGA 1091
 DB 356 TCGCGTTTCTCGCGTGAAGGTACGCGACGAATGCGCAGGCCTA 401

RESULT 11
 ABS58076
 ID ABS58076 standard; DNA; 1248 BP.
 XX
 AC ABS58076;
 XX
 DT 05-FEB-2003 (first entry)
 XX
 DE Mesorhizobium loti btaA gene, Ml-btaA.
 XX
 KW Betaine lipid production; betaine lipid compound; DGTS;
 KW diacylglycerol-0-4'-(N,N,N-trimethyl) homoserine; agricultural industry;
 KW phosphate-containing fertilizer; transgenic; btaA; Ml-btaA; gene; ds.
 XX
 OS Mesorhizobium loti.
 XX
 PN WO200283844-A2.
 XX
 PD 24-OCT-2002.
 XX
 XX 09-APR-2002; 2002WO-US011134.
 PF
 XX 13-APR-2001; 2001US-0283812P.
 PR
 PR 08-APR-2002; 2002US-00283812.
 XX
 XX (UNMS) UNIV MICHIGAN STATE.
 PA
 XX Benning C, Riekhof W, Klug R;
 PI
 XX WPI; 2003-058632/05.
 DR
 XX
 PT New composition comprising an isolated and purified DNA molecule, useful
 PT for producing Betaine lipids, e.g. Diacylglycerol-0-4'-(N,N,N-trimethyl)
 PT homoserine (DGTS) for agricultural applications.
 XX
 XX Claim 1; Fig 22; 109pp; English.
 PS
 XX The present invention relates to compositions and methods for producing
 CC betaine lipids. The method comprises the expression of recombinant
 CC enzymes (e.g. from Rhodospirillum rubrum) in host cells such as
 CC bacteria, yeast, and plants to produce betaine lipid compounds including
 CC diacylglycerol-0-4'-(N,N,N-trimethyl) homoserine (DGTS). The methods and
 CC compositions of the invention are useful in agricultural applications,
 CC such that the amount of phosphate-containing fertilizer required for the
 CC growth of a particular plant is decreased. The polynucleotide sequences
 CC encoding the recombinant enzymes may be used to produce vectors which can
 CC be used to produce transgenic plants. The present sequence represents the
 CC R. sphaeroides btaA gene homologue from Mesorhizobium loti, Ml-btaA
 XX
 XX Sequence 1248 BP; 224 A; 441 C; 345 G; 238 T; 0 U; 0 Other;
 SQ
 Query Match 3.6%; Score 69.6; DB 8; Length 1248;
 Best Local Similarity 45.0%; Pred. No. 0.0017;
 Matches 500; Conservative 0; Mismatches 579; Indels 33; Gaps 5;
 QY 704 TCCCTACGTCGTCGTCGCGGCCCTTACTACGTGTGATTTGGCGCTGCCACGCG 763
 DB 11 TCTCTCGGATCTGTTTTCGCGCGGCAAGAGTTGAAAGCGCTCTACAGAAC 70
 QY 764 TTGGCCACGCTCTGACGAGGAGCGGTGGAGCGCGGCCCATGTTCCCGCCACCTTCC 823
 DB 71 GCGGCTTTCCAAAGCGGCATCTCGAGCGGCTGTTCCGCTTCTCTTTTCCGCGCTCG 130
 QY 824 TGTACACGAGTGTGGAGGACCCCGAGCGGATATGGAGGTGATGAGATCAACCCCA 883
 DB 131 TCTATCCGAGATCTGGAGAGACCCCGATGTCACATGGAGGCCATCAGCTTGGTCAGG 190
 QY 884 AGACACGCTGTGACCTGACTAGCGCGGCTGCAATGCCCTGAACTCTGTGTGTCAGG 943
 DB 191 GCATCGCATCGTCACATCGCTTCCGCGGCTTGCACATCTCTACCTACCTACCGTT 250
 QY 944 GGGCCGCGCAGGTGGTGTGCTGGAGCTGCAACCCCGCGAGTGGCGCTTCTGGAGCTGA 1003

Db 251 CGCGGACGGATCGAGCGCGTCTGACCTCAACGCGCGCCCAATCGCGCTGAACCGCATGA 310
Qy 1004 AGAAGTGGCGCAATTCAGCAGCTGAGTTTGAGGACG---TGTGGCAGCTTTCGGCGAGG 1060
Db 311 AGCTGGAGGCGGTGCGCGCTGCTCCCTCGCAGGCGATCTGTTCCGCTTTTTCGGCGCG 370
Qy 1061 GCGTGCAACCGCGGATGAGAGCTGTACAGAGAAAGCTGGCGCCCTTCTCTGTCGAAA 1120
Db 371 CGGACACACGACCAATTCGCAAGCCTATGACCGCTTTATTTGGCGCGCATCTCGATCCGG 430
Qy 1121 CCAGCCACAACTTCTGTCTCAAGCGCTCTGTGTTCTTCCAGCAGCGGCTGTACTACGAG 1180
Db 431 TCAGCCGCACTATTGGAGCGCGCACTGGGTGTCGCCGCGCATCCCGTCTTCG 490
Qy 1181 CGGCGATGGCAAGCTGTGTCTGAGTGTCTGAGTGCCTTGGCGCGCTGTGGAGCTGGCA 1240
Db 491 ACCGCAATTTCTACGAGC---GGCCTGTCTGCGCTGTTTCATCGCCATGGCCATCGCA 547
Qy 1241 AGACCGTCAAGCGCTCGCAAGCGCGCCCAATGAGGAGCAGCGGCTGTGTGGGACA 1300
Db 548 CGGCGAAATTTCTTGGCGTCAACCGCGCCCACTGATGGAAGCAGGAAATATCGGCGAGC 607
Qy 1301 GCAACATGCTCATCTCTGTGTAAGAAAGCGGCCCAAGCGCTGTGTGCTGTCTGTC 1360
Db 608 AGCGCGCTTCTTCAACGAG---AGCTGGCGCGGTCTTCAACAAG 652
Qy 1361 AGTTGCGAGCTGTGTCTTCAACAAGCGCGTGTGTGTGTTGCGCGCGGTGCGCG 1420
Db 653 AGCTTTTGAATGGCGACCTCGCTGAAGCGCTCGCTGTTTCGGCTCGGCAATTCGCGCG 712
Qy 1421 GCAAGCATGAGCGCTGATCAAGCGGAGCGGATCCCATTTGAGAACTATCATCGCGGCA 1480
Db 713 CGCAGTACGATTCCTGATCACCTCAGCGCAGCGGCACATGGCGCAGCGTTCTGAAGGCC 772
Qy 1481 CCATGACGCGGTGGGGAATCTCGCAGCTGCGCAGCAGCACTTCTTACTACACT 1540
Db 773 GCGTGGAAGCTCGCTCGGATTTTCCCTTGGAACAATTTATTCGCTGCGAGCTT 832
Qy 1541 GCCTCACCGCAAGTTCTCGCGGCACAACTGC-----CCCACTACTCGCGGAGCGG 1594
Db 833 TTGCGCGCGCTATCCAAATCCGCTGAGCGCGCTGCGCGCTATCTGGAAGCAGA 892
Qy 1595 CTTTGGCACCT-----CAAGATGGCGTGTGAGCAACCTGACGCTTCCACCACT 1648
Db 893 ACTACGAAACCATCCGCGCAATATCGACGCGTGCCTCCACCATGCCAATCTGATCG 952
Qy 1649 TCTTATGAGGAGCTCAAGCGCGCCTTACACCAAGGTGATTCTGATGACCACTGG 1708
Db 953 AATTCCTCGCGGAGGAGCGCGGACCTGATCGCTTATCTCTGCTGATGGCGAGG 1012
Qy 1709 ACTGGCTGATATGCCCCGTGGCCAAAGAGCTGGCGGAGTGCCTGGCCAAAGCAGTTTGGCG 1768
Db 1013 ACTGGATGACGATGACCATCAACGCGCTGTGTTGCGGAATCAGCGCGACCGCTCG 1072
Qy 1769 CGGCGGCGATGCTGATCTGCGCTCGCTCGCTCC 1800
Db 1073 CAGCGCGCGCTCATCTTCCGCAACCGCGCC 1104

RESULT 12

ADW72725
ID ADW72725 standard; DNA; 1248 BP.

XX

AC

XX

DT

21-APR-2005 (first entry)

XX

DE Mesorhizobium loti btaA gene.

XX

KW Phospholipid synthesis; transgenic plant; fertilizer; ss; gene;

KW betaine lipid.

XX

OS Mesorhizobium loti.

XX Key Location/Qualifiers
FH CDS 1. .1248
PT /*tag= a
FT /product= "BtaA protein"
XX

W02005009115-A2.

XX

PD 03-FEB-2005.

XX

PF 15-JUL-2004; 2004W0-US022789.

XX

PR 16-JUL-2003; 2003US-00620914.

XX

PA (UNMS) UNIV MICHIGAN STATE.

XX

PI Benning C, Riekhof W;

XX

DR WPI; 2005-112975/12.

XX

DR P-PSDB; ADW72744.

XX

PT New composition comprising a purified DNA having an oligonucleotide

XX

PT sequence encoding a protein, useful in producing Betaine lipid compounds

XX

PT e.g., Diacylglycerol-O-4'-(N,N,N-trimethyl) homoserine (DGTS).

XX

PS Disclosure; SEQ ID NO 22; 147pp; English.

XX

XX

CC The invention relates to a new composition comprising a purified DNA

CC

CC having an oligonucleotide sequence appearing as ADW72747 (Chlamydomonas

CC

CC reinhardtii BtaI cDNA) or ADW72752 (Neurospora crassa BtaI coding region,

CC

CC encoding BtaI proteins which are DGTS (diacylglycerol-O-4'-(N,N,N-

CC

CC trimethyl)homoserine) synthetic enzymes. BtaI (and BtaA and BtaB) enzymes

CC

CC allow the replacement of phospholipids with non-phosphorus containing

CC

CC lipids in the cell membrane. A transgenic plant expressing the enzymes

CC

CC would have a reduced need for phosphate-containing fertilizer. Also

CC

CC included are an RNA transcribed from the purified DNA, antibodies

CC

CC produced from the protein, a vector comprising the DNA or nucleic acid, a

CC

CC host cell comprising the vector, transgenic plants comprising the vector,

CC

CC and a protein translated from the RNA or encoded by the nucleic acid. The

CC

CC composition is useful in producing Betaine lipid compounds e.g.,

CC

CC diacylglycerol-O-4'-(N,N,N-trimethyl) homoserine (DGTS). The composition

CC

CC of the invention reduces the amount of phosphate fertilizer needed for

CC

CC the optimal growth of crop plants. The present sequence encodes a betaine

CC

CC lipid synthetic enzyme.

XX

SQ Sequence 1248 BP; 224 A; 441 C; 345 G; 238 T; 0 U; 0 Other;

Query Match 3.6%; Score 69.6; DB 14; Length 1248;

Best Local Similarity 45.0%; Pred. No. 0.0017;

Matches 500; Conservative 0; Mismatches 579; Indels 33; Gaps 5;

Qy 704 TCCCTTACGTGCGTGGCTGCGCGCCCTTACTACTGCTGTGATGCGCGCTGCCCCAGCG 763

Db 11 TCTCTCTCGATCTGGTTTTCGCGCGGCAAGAGTTGGAAGGCGCTTACCAAGACC 70

Qy 764 TTGGCCACCCCTGACAGGAGCGCGTGGAGCGCGGCCCATGTTCCCGCCCACTTCC 823

Db 71 GCGCGCTTTCCAAAGCGCGCATCTCCGAGCGGCTGTTGCGCTTCTCTGTTTTCGCGCTCG 130

Qy 824 TGTACAGCAGTCTGGGAGGAGCCCGAGCCGATATGAGGTGATGGAGATCAACCCCA 883

Db 131 TCTATCCGAGATCTGGGAAGACCCCGATGTCACATGGAGGCGCATGCGCTTGGTCAGG 190

Qy 884 AGGACACGCTGTGACCTGACTAGCGCGGCTGCAATGCCCTGAACTGCTGCTGTGCGAGG 943

Db 191 GCCATCGCATGTCACAAATCGCTTCCGCGGCTGCAACATCTCGCTTACCTACCGGTT 250

Qy 944 GGGCCCGCCAGGTGGTGTGCTGGAGTGCACACCCCGCGAGTGGCGCTTCTGGAGCTGA 1003

Db 251 CGCGGACCGATCGACGCGCTGCACTCAACGCGGCCACACATCGCGCTGAACCGCATGA 310

Qy 1004 AGAAGTGGCCATTCAGCAGCTGGAGTTCAGGACG---TGTGGCAGCTGTTGGCGAGG 1060

```
Db 311 AGCTGGAGGCGGTGCGCGCTGTCCTCGCAGGCGGATCTGTTCCGCTTTTTCGGCGCG 370
Qy 1061 GCGTGACCCCGGATTTGAGAGCTGTACAGAAAGCTGGCGCCCTTCTCTGCGCAA 1120
Db 371 CCAGACACAGCCCAATTCGCAAGCCTATGACCGCTTTATTTGGCGCGCATCTCGATCCGG 430
Qy 1121 CCAGCCACAATTTCTGTCCTCAAGCGCTCTGCTGTTACTTCCAGCAGCGGCTGTACTACCAGG 1180
Db 431 TCAGCGCCCACTATTGGAGCGCGCACTGGGTGTCGCCGCGCATCGCGCTTTCG 490
Qy 1181 GCGGCATGGCAAGCTGTGCTGGGTGCTGCAAGTCCCTGGCGCGCTGGTCTGGGACTGGGCA 1240
Db 491 ACCGCAATTTCTACGAGC---GGCCTGCTCGGCTGTTTCACTCGCCATGGGCAATCGCA 547
Qy 1241 AGACCGTCAAGCGCTCGCAAGCGCCCAATGAGGAGCAGCGCCGCTCTGTGGACA 1300
Db 548 CGGCGAAATTTCTTGGCGCTCAACCGCGCCCACTGATGGAAGCAGCAATATCGCGGAGC 607
Qy 1301 GCAACATGCTCATCCACTTCTGTAAGAAAGCGGCCCAAGCGCTGGTGTGCTGTGTCTCA 1360
Db 608 AGCGCCGCTTCTTCAACAGG-----AGCTGGCGCGGTCTTTCGACAGA 652
Qy 1361 AGTTCTGAGCGCTGCTCTTCAACAAAGCGCTGCTGTGTTTCGGCGCGCGGTGCGCG 1420
Db 653 AGCTTTTGAATGGCGACCTCGCGTAAGCGCTCGCTGTTTCGGCTCGGCATTCGCGCG 712
Qy 1421 GCAAGAGTACGCGCTGATCAAGCGGAGCGGATCCCATTTGAGAACTACATCGCGGCA 1480
Db 713 CGCAGTACGATTCCTGATCACTCAGCGCAGCGCACCATGGCCAGCGTTCGAAGGCC 772
Qy 1481 CCATGGAGCGGTGGCGGAGACTCGCACGTGGCAAGCAGAACTACTTCTACTACAAT 1540
Db 773 GGCTGGAAGCTCGCTCGGATTTCCCTGGAAACAATTTATTCGCTGGCAGGCTT 832
Qy 1541 GCCTCAGCGGAGTTCCTTGGCGCAAACTGC-----CCACCTACTCTGCGAGCGG 1594
Db 833 TTGCCCGCGCTATCCAAATCCCGGTGAGCGCGCTGCGCGCTATCTGGAAGCAGA 892
Qy 1595 CTTTGGCCACCT-----CAAGAGTGGGTGGTGGCAACCTGACCGTCTCCACCACT 1648
Db 893 ACTACGAACCATCCGCGCAATATCGACCGCTGCGCATCCACCATGCAATCTGATCG 952
Qy 1649 TCTTCATGAGGAGCTCAAGCGCGCACCTACACCAAGGTGATTCGATGACCACTGG 1708
Db 953 AATTCTCGCGGCAAGACGCGGCACTCGATCGCTTCACTCTGCTGATGCGCAGG 1012
Qy 1709 ACTGGCTGGATATGCCGTGGCCCAAGCAGTGGCGCGAGTGCCTGGCCCAAGCAGTTGCGC 1768
Db 1013 ACTGATGACCGATGACCACTCAACGCGCTGTGTCGGAATCAGCGCGCACGCGCTCG 1072
Qy 1769 CGGCGCGCATGCTATCTGGGCTTCGCGCTCC 1800
Db 1073 CAGCGCGCGCTATCTTCGCAACCGCGCGC 1104
```

RESULT 13

AAH4047

ID AAH4047 standard; DNA; 1092 BP.

XX AC

XX AC

XX AC

DT 10-SEP-2001 (first entry)

XX Streptomyces sp. CL190 mevalonate pathway orfD DNA SEQ ID NO:5.

XX Streptomyces sp. CL190; mevalonate pathway; actinomycete; cardiant;
XX isoprenoid compound; osteopathic; cytosolic; ubiquitonic; vitamin K2;
XX carotenoid; heart disease; osteoporosis; cancer; drug; health food; ds.

XX Streptomyces sp. CL190.

XX Key

XX Location/Qualifiers

XX 1..1092

XX CDS

/*tag= a
/product= "orfD"

WO200142476-A1.

14-JUN-2001.

06-DEC-2000; 2000WO-JP008620.

08-DEC-1999; 99JP-00348375.

(SETO/) SETO H.

(KUZU/) KUZUYAMA T.

Seto H, Kuzuyama T, Takahashi S, Takagi M;

WPI; 2001-381696/40.

P-PSDB; AAB99729.

Actinomycetes-originated genes of enzymes participating in mevalonate pathway, applicable in producing e.g. ubiquitonic, vitamin K2 and carotenoids for treatment of heart diseases, osteoporosis and cancer in drug and health food.

Claim 4; Page 50-52; 75pp; Japanese.

The sequence given in AAH4043 represents a DNA sequence isolated from Streptomyces sp. CL190, containing a 6798 base pairs (SI), which encodes the whole enzyme necessary for functioning the mevalonate pathway. The sequence encodes protein sequences, designated orfA to E and hmgr, which are used in the mevalonate pathway. The proteins and polynucleotide sequences encoding them have cardiant, osteopathic and cytostatic activities. The genes are applicable in producing e.g. ubiquitonic, vitamin K2 and carotenoids which can be used in the treatment of heart diseases, osteoporosis and cancer in drugs and health foods. The present sequence encodes the orfD protein from the present invention

SQ Sequence 1092 BP; 181 A; 430 C; 340 G; 141 T; 0 U; 0 Other;

Query Match 3.6%; Score 69.4; DB 4; Length 1092;

Best Local Similarity 45.3%; Pred. No. 0.0013;

Matches 253; Conservative 0; Mismatches 306; Indels 0; Gaps 0;

Qy 1344 GGTGTGCTCTTCGTCAAGTTCTGAGCCTGTGCTCTTCAACAAGCCCTGTGTGTT 1403

Db 501 GTTCCGCGATCGAAGATCGCGGCGCTCGACATCCCGTGTCTCAAGAGGT 560

Qy 1404 CGCGCGCGCTCCCGGCAAGCAGTACGCGCTGATCAAGCGGAGCGGATCCCATTTGA 1463

Db 561 CGGCAACGGCTGAGCGCGGAGACCATCTGCTGCTCGCGGCTCCAGGCGGC 620

Qy 1464 GAATACATCGCGCGCACCATGAGCGGCTGGCGGAGACTCGCATGTCGCAAGCAGAA 1523

Db 621 GGACGTGAGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGAGCT 680

Qy 1524 CTACTTCTACTACAATGCTCTACCGGCAAGTTCTTCTGCGGCAAACTGCGCCACCTACCT 1583

Db 681 CGCGGACTACCGTTCTGACCGCTGGGGGCGAGTCCACCGCGCTGCTGTGGAGCG 740

Qy 1584 GCGGAGCGCGCTTCGCCACCTCAAGAGTGGCGGTGGCAAACTGACCGTCTTCCAC 1643

Db 741 CCAGGACATCTCCCTGCGCGCTCTCGCTCGCGGCTGTGCTCACCGCTCGACGTTGT 800

Qy 1644 CAACTTCTTATGAGGAGCTCAAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1703

Db 801 CGCGCGCTCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 860

Qy 1704 CGTGGACTGCTGATATGCCGTGGCAACAGAGTGGCGGAGTGGCTGTGGCCAGCAGGT 1763

Db 861 GATGAGCAGCGGCTCGACCGGCTGATCAGAGCTCACGACCTGGCTGGACGAGTGGC 920

Qy 1764 TGCGCGCGGCGGATCGTATCTGCGGCTCGCGCTCGCGCTCGCGCTCGCGCTCGCG 1823

Db 1764 TGCGCGCGGCGGATCGTATCTGCGGCTCGCGCTCGCGCTCGCGCTCGCGCTCGCG 1823

Db 921 GCGCTGCGAGACCATGCTCGGCGCGCGCACCCCGCGCGACCTCAACCGCTGCGAGCTGCT 980
QY 1824 GATCCAGAACGCGGCTTCGACGTGCGCTGCATCCGCGCGCGCACTCAGGGCTACATGGA 1883
Db 981 GCTCCAGCGGAGCTGCGTGAATCTTCTGCGCGGACCGGGGCGATCGACACGCGCGCTCGC 1040
QY 1884 CCGGCTCAACATGTACAGC 1902
Db 1041 CCAGCGCTCCAGTCCATC 1059

RESULT 14

ABD16186

ID ABD16186 standard; DNA; 3003 BP.

AC ABD16186;

DT 29-JUL-2004 (first entry)

XX Pseudomonas aeruginosa polynucleotide #14790.

KW Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;

KW antibacterial.

OS Pseudomonas aeruginosa.

XX US6551795-B1.

XX 22-APR-2003.

XX 18-FEB-1999; 99US-00252991.

XX 18-FEB-1998; 98US-0074788P.

PR 27-JUL-1998; 98US-0094190P.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX Rubenfield MJ, Nolling J, Deloughery C, Bush D;

XX WPI; 2003-615309/58.

DR P-PSDB; ABO82615.

Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide, useful as molecular targets for diagnostics, prophylaxis and treatment of pathological conditions resulting from bacterial infection.

PS Disclosure; SEQ ID NO 14790; 455pp; English.

The invention relates to Pseudomonas aeruginosa polypeptides and the polynucleotides encoding them. The sequences are useful in diagnosis and therapy of pathological conditions, as molecular targets for diagnostics, prophylaxis and treatment of pathological conditions resulting from a bacterial infection, for evaluating a compound, such as a polypeptide, for the ability to bind a P. aeruginosa nucleic acid, as components of effective antibacterial targets, as targets for antibacterial drugs, including anti-P. aeruginosa drugs, as templates for recombinant production of P. aeruginosa-derived peptides or polypeptides, as target components for diagnosis and/or treatment of P. aeruginosa-caused infection, and in detection of P. aeruginosa sequences or other sequences of Pseudomonas species using biochip technology. Sequences ABD01397-ABD17967 represent P. aeruginosa polynucleotides of the invention. Note: The sequence data for this patent did not form part of the printed specification but was obtained in electronic format from USPTO at seqdata.uspto.gov/sequence.html

SQ Sequence 3003 BP; 488 A; 1076 C; 959 G; 480 T; 0 U; 0 Other;

Query Match 3.6%; Score 69.4; DB 11; Length 3003;

Best Local Similarity 43.8%; Pred. No. 0.002;

Matches 454; Conservative 0; Mismatches 571; Indels 12; Gaps 3;

QY 521 TGGGCGCTTCCGACTTCTAGTGGCGGCGGACGACCTGCGCCCTGCGCGCATGCCCT 580

RESULT 15

ABD15916

Db 1965 TCGGCGTTCGACGGCAGCAACCCCGACCCCTCAACACCGGCGCCCTGTGTGATCAACCTCAAGC 2024
QY 581 GGTTCGCGCGCTTTCTTCTTCTGCGCATCGATCTTTCGACATCGAACAATTTGATCATCGGCCCG 640
Db 2025 CGCACAGCAGCGCGAGCTCACCGCAGCGAGTGATCCAGCGCTTCAGCGCCGAACTCG 2084
QY 641 AGCGCGCGCTTACCTTGGAGCAGAGCTGGAGCGCTGTGGAGAGCAGAACACCCAGGGTT 700
Db 2085 ACCACCTGCGCGGATCAAGCTGTATATGACGCGGTGACGAGACCTGACCATCGAGACC 2144
QY 701 CGATCCCTACGTGCGCTGCGCGCCCTTACTACGTGTGTGATTTGGCGCTGCGCCA 760
Db 2145 GGTTCGCGCGCACCGAGTACAGTTTCACTTGCAGGACGCCGACCGAGCTGCTCGCG 2204
QY 761 GCGTTGGCCACGCCCTGCGACGAGAGCGGTGGAGCGCGGCCCATGTTTCCCGCCACCT 820
Db 2205 AGTGGGTGCGGAGCTGGTGGCGCGCTGCAGGAGTTGCGCGAGCTCGCCGACGTGCGCA 2264
QY 821 TCCTGTACAGCAGTCTGTGGAGGACCCCGAGCGGATATGGAGGTGATGGAGATCAACC 880
Db 2265 GCGACTGGCAGGACAAAGGGCTTGCAGGCTTACCTGAACATCGACCGCG-----ACACCG 2318
QY 881 CCAAGGACACGCTGTGACCTGACTAGCGCGGCTGCAATGCCCTGAACCTGCTGTGTGC 940
Db 2319 CTTGCGCTTGGCGTGAAGCTCTCCGACATCGACAGCTGTCTTACACGCTTTCGCC 2378
QY 941 AGGGGCGCGCGCAGGTGTGTGCTGCGTGTGACTGTCAACCCCGCGAGTGGCGCTTCTGAGC 1000
Db 2379 AGCGGTGATCTCGACCATCTTTCACCCAGGCCACCCAGTACCGGTGGTGTGAGGTGG 2438
QY 1001 TGAAGAGGTGGCCATTTCAGCAGCTGGAGTTTGAAGACGTGTGGCAGCTGTTCGGCGAGG 1060
Db 2439 CGCGCAGTTTCCAGCTTCGGCGCGCGCAGGCGCTTGAAGCAGCTCTACGTGCGCTCCAGCGACG 2498
QY 1061 GCGTGACCGCGCATTTGAGGAGCTGTACGAGAAGAGCTGGCGCGCTCTCTGTGCAAA 1120
Db 2499 GCACCGAGTGGCGCTGTGAGCTTGGCGAGGTGGAGAGCGGCGATACCTGCTGCGCA 2558
QY 1121 CCAGCCACAACCTTCTGGTCCAAAGCGCTCTGTGACTTTCAGCAGCGGCTGTACTACAGG 1180
Db 2559 TCAACCATATCGCCAGTTTC---CCCTCGCGGACCCCTGTCTTCAACCTTGGCCAAAGGTT 2615
QY 1181 GCGGCATGGGCAAGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTG 1240
Db 2616 ACTCCCTGCGCGAGCGGCTCGAGCGGATCGGTGCGGTTCGAGGCGCAGCTGTGAGCTGCGCG 2675
QY 1241 AGACCGTCAAGCGCTTCGCCCAAGCGCCCAATGAGGAGCAGCGCGCTCTGTGGGACA 1300
Db 2676 TGAAGCATGAGGCGAGCTTCGCGCGCGCGCGCTGCGCTTTCGAGGCGCTGCTGTGCAACA 2735
QY 1301 GCAACATGCTCATTC---ACTTCTGTGAAGAACGGGCGCCCAAGCGCTGTGTGTGCTGTGCTG 1357
Db 2736 CGCTGCTGTGATCTCGCTCGCTCGGTGCTGACCATGTACATCGCTCTGGGCATCTCTACG 2795
QY 1358 TCAAGTTCTGAGCTGTGCTCTTCAACAGCGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTG 1417
Db 2796 AGAGCTTTCATTCATCGGCTGACCATCTCTCGACCCCTGCGCTGCGCGCGGCTGCGCGCGC 2855
QY 1418 CCGGCAAGCAGTACGCGCTGATCAAGCGGCGGCGCATCCCATTTGAGAACTACATCGCGC 1477
Db 2856 TGTGCGCTGTGCTGTGCGCGGCGGCGAGAGATCGGCATCTGTGGCGATCATCGGCATCATCC 2915
QY 1478 GCACCATGACGCGCTGTGCGGAGAACTCGCAGTGTGCGCAAGCAGAACTACTTCTACTACA 1537
Db 2916 TGCTGATCGCATCGTCAAGAGAACCGCATCATGATGATGCTGCTTCCGCTTCGAGCGCG 2975
QY 1538 ACTGCTCTACCGGCAAG 1554
Db 2976 AGCGCAACGAAGCGAAG 2992

ID	ABD15936	standard; DNA; 3132 BP.
XX	ABD15936;	
XX	AC	
XX	AC	
XX	DT	29-JUL-2004 (first entry)
XX	XX	
DE	Pseudomonas aeruginosa	polynucleotide #14540.
XX	XX	
XX	Bacterial infection; gene; ds;	Pseudomonas aeruginosa infection; antibacterial.
KW	XX	
KW	XX	
XX	OS	Pseudomonas aeruginosa.
XX	OS	
PN	US6551795-B1.	
XX	XX	
PD	22-APR-2003.	
XX	XX	
PF	18-FEB-1999;	99US-00252991.
XX	XX	
PR	18-FEB-1998;	98US-0074788P.
PR	27-JUL-1998;	98US-0094190P.
XX	XX	
XX	(GENO-)	GENOME THERAPEUTICS CORP.
PA	XX	
XX	XX	
PI	Rubenfield MJ, Nolling J, Deloughery C, Bush D;	
XX	XX	
XX	XX	
DR	WPI; 2003-615309/58.	
DR	P-PSDB; ABO82365.	
XX	XX	
PT	Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,	
PT	useful as molecular targets for diagnostics, prophylaxis and treatment of	
PT	pathological conditions resulting from bacterial infection.	
XX	XX	
PS	Disclosure; SEQ ID NO 14540; 455pp; English.	
XX	XX	
CC	The invention relates to Pseudomonas aeruginosa polypeptides and the	
CC	polynucleotides encoding them. The sequences are useful in diagnosis and	
CC	therapy of pathological conditions, as molecular targets for diagnostics,	
CC	prophylaxis and treatment of pathological conditions resulting from a	
CC	bacterial infection, for evaluating a compound, such as a polypeptide,	
CC	for the ability to bind a P. aeruginosa nucleic acid, as components of	
CC	effective antibacterial targets, as targets for antibacterial drugs,	
CC	including anti-P. aeruginosa drugs, as templates for recombinant	
CC	production of P. aeruginosa-derived peptides or polypeptides, as target	
CC	components for diagnosis and/or treatment of P. aeruginosa-caused	
CC	infection, and in detection of P. aeruginosa sequences or other sequences	
CC	of Pseudomonas species using biochip technology. Sequences ABD01397-	
CC	ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:	
CC	The sequence data for this patent did not form part of the printed	
CC	specification but was obtained in electronic format from USPTO at	
CC	seqdata.uspto.gov/sequence.html	
XX	XX	
SQ	Sequence 3132 BP; 496 A; 1133 C; 993 G; 510 T; 0 U; 0 Other;	
Query Match	3.6%;	Score 69.4; DB 11; Length 3132;
Best Local Similarity	43.8%;	Pred. No. 0.0021;
Matches	454; Conservative	0; Mismatches 571; Indels 12; Gaps 3;
QY	521	TGGCGGTTCGGACTTCTACGTGAGCGGCAAGTACGACCTGCCCTCGGCCACATGCCCT 580
Db	1799	TCGGGCTCGACGGCAGCAACCCGACCTCAACACCGCGCGCTGCTGATCAACCTCAAGC 1858
QY	581	GGTCGCGCCGTTTCTTCCTGGCGATCGATCTTCGACATCGACACATTGACATCGGCCCG 640
Db	1859	CGCACAGCAGCGCAGCGTCAACCGCCAGCGAATGATTCAGCGCTTCGACGCCGAACTCG 1918
QY	641	AGCGCCCGCGCTTACTCTCGAGCAGAAGCTGGAGCGCGTGTGGGAGCAGAAACACCCAGGTT 700
Db	1919	ACCACTGCCCGGATCAAGCTGTACATGCAGCGCGTGCAGGACCTGACCATCGAGGACC 1978
QY	701	CGATCCCTTACGTGCGTGGTGGCGGCCCTTACTACGTGTGATTTGGCGGCTGCCTCA 760
Db	1979	GGGTGCGCCCGCACCACGTACAGTTTCACTTGCAGGACGCGGACCCCGGACGTGCTCGCCG 2038

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 14, 2006, 01:29:45 ; Search time 49.8 Seconds
(without alignments)

5516.721 Million cell updates/sec

Title: US-10-620-914-44

Perfect score: 3619

Sequence: 1 acggggctcgggtcgtgacgg.....gcccgaagaagacaactaa 1947

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 4332886

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlh
-O=/abes/ABSSWEB_spool/US10620914/runat_13032006_101925_5419/app_query_faeta_1
-DB=UniProt -QFMT=fastan -SUFFIX=n2p.rup -MINMATCH=0.1 -LOOCL=0 -LOOEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blom62 -TRANS=human40.cdi -LIST=45
-LOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abes05h
-USER=US10620914 @CGN 1.1.192 @runat_13032006_101925_5419 -NCPU=6 -ICPU=3
-NO_WMAP -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : UniProt 05.80.*

1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3441	95.1	666	2 Q6DN05	CHLRE
2	953	26.3	835	2 Q4IPW6	GIBZE
3	924	25.5	944	2 Q4PFP5	USTWA
4	910	25.1	741	2 Q6CEB3	YARLI
5	891	24.6	752	2 Q59TS3	CANAL
6	890.5	24.6	790	2 Q6CJ18	KLJLA
7	881.5	24.4	832	2 Q52GF6	MAGGR
8	881	24.3	820	2 Q55J70	CRYNE
9	864.5	23.9	763	2 Q6BK65	DEBHA
10	850.5	23.5	908	2 Q7SGY5	NEUCR
11	845.5	23.4	797	2 Q5KGX9	CRYNE
12	806.5	22.3	831	2 Q4X175	ASPFU
13	752	20.8	790	2 Q5B0L6	EMENI
14	463	12.8	415	2 Q7UYV8	RHOBA
15	413.5	11.4	767	2 Q4NR00	_9DELT
16	409.5	11.3	251	2 Q7UYV9	_RHOBA

C	17	401	11.1	2268	2	Q4NRAS_9DELT	Q4nr5 anaeromyxob
	18	380.5	10.5	407	2	Q6NEV1_RHOSH	Q6nbv1 rhodopsedo
	19	379.5	10.5	416	2	Q93TQ1_RHOSH	Q93tq1 rhodobacter
	20	368	10.2	767	2	Q4NR00_9DELT	Q4nr00 anaeromyxob
	21	364	10.1	964	2	Q4NOB4_9DELT	Q4nqb4 anaeromyxob
	22	357	9.9	2147	2	Q4NUU3_9DELT	Q4nuu3 anaeromyxob
	23	354.5	9.8	1122	2	Q4NQW1_9DELT	Q4nqw1 anaeromyxob
	24	354	9.8	1286	2	Q9VR49_DROME	Q9vr49 drosophila
	25	353	9.8	1549	2	Q4NU10_9DELT	Q4nu10 anaeromyxob
	26	349	9.6	905	2	Q4NTZ7_9DELT	Q4ntz7 anaeromyxob
	27	345.5	9.5	608	2	Q9Q5K9_CHV12	Q9q5k9 cercopithec
	28	340	9.4	761	2	Q4NT26_9DELT	Q4nt26 anaeromyxob
	29	338	9.3	1549	2	Q4NU10_9DELT	Q4nu10 anaeromyxob
	30	336.5	9.3	964	2	Q4NRUS_9DELT	Q4nrus anaeromyxob
	31	335	9.3	863	2	Q4NT95_9DELT	Q4nt95 anaeromyxob
	32	333	9.2	2268	2	Q4NRAS_9DELT	Q4nr5 anaeromyxob
	33	328	9.1	964	2	Q4NOB4_9DELT	Q4nqb4 anaeromyxob
	34	327	9.0	890	2	Q8UZE1_9GAMA	Q8uz1 cercopithec
	35	326.5	9.0	416	2	Q8UDK6_AGRFS	Q8udk6 agrobacteri
	36	326	9.0	695	2	Q4NX38_9DELT	Q4nx38 anaeromyxob
	37	323.5	8.9	660	1	YHL1_EBV	P03181 Epstein-Bar
	38	323.5	8.9	660	1	Q777A3_9GAMA	Q777a3 human herpe
	39	323	8.9	993	2	Q4NMX3_9DELT	Q4nm3 anaeromyxob
	40	322.5	8.9	706	2	Q41972_MHV68	Q41972 murid herpe
	41	322.5	8.9	880	2	Q4NU98_9DELT	Q4nu98 anaeromyxob
	42	321.5	8.9	774	2	Q41971_MHV68	Q41971 murid herpe
	43	321	8.9	432	2	Q98K98_RHILIO	Q98k98 rhizobium 1
	44	318	8.8	2013	2	Q4NWX4_9DELT	Q4nwx4 anaeromyxob
	45	316	8.7	1309	2	Q4NU17_9DELT	Q4nu17 anaeromyxob

ALIGNMENTS

RESULT 1

ID	Q6DN05	CHLRE	PRELIMINARY;	PRT;	666	AA.
AC	Q6DN05;					
DT	25-OCT-2004	(TREMBLrel. 28, Created)				
DT	25-OCT-2004	(TREMBLrel. 28, Last sequence update)				
DT	25-OCT-2004	(TREMBLrel. 28, Last annotation update)				
DE	Betaine lipid synthase.					
GN	Name=BTL;					
OS	Chlamydomonas reinhardtii.					
OC	Eukaryota; Viridiplantae;					
OC	Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.					
OX	NCBI_TaxID=3055;					
RN	[1]					
RP	NUCLEOTIDE SEQUENCE.					
RC	STRAIN=CC125;					
RX	PubMed=15701786; DOI=10.1128/EC.4.2.242-252.2005;					
RA	Riekhof W.R., Sears B.B., Benning C.;					
RT	"Annotation of Genes Involved in Glycerolipid Biosynthesis in					
RT	Chlamydomonas reinhardtii: Discovery of the Betaine Lipid Synthase					
RT	BTAICr.;"					
RL	Eukaryot. Cell 4:242-252(2005).					
DR	EMBL; AY656806; AAT72342.1; -; mRNA.					
DR	GO; GO:0008757; F:S-adenosylmethionine-dependent methyltransf. . . ; IEA.					
DR	GO; GO:0016740; F:transferase activity; IEA.					
DR	InterPro: IPR002114; Hpr SerP_S.					
DR	InterPro: IPR000051; SAM_bind_S.					
DR	PROSITE; PS00589; PTS_HPR_SER; UNKNOWN_1.					
KW	Methyltransferase; Transferase.					
SQ	SEQUENCE 666 AA; 75780 MW; 486A6F5718682284 CRC64;					

Alignment Scores:						
Pred. No.:	1,14e-169	Length:	665			
Score:	3441.00	Matches:	647			
Percent Similarity:	97.3%	Conservative:	1			
Best Local Similarity:	97.1%	Mismatches:	0			
Query Match:	95.1%	Indels:	18			
DB:	2	Gaps:	1			

US-10-620-914-44 (1-1947) x Q6DN05_CHLRE (1-666)

QY 1 ATGGGGTGGGTGCTGAGCGCGCGCTCGAGCTTACACCAAGAACTTCTCCCTGGAG 60
 Db 1 MeGlySerGlyArgAspGlyArgProAlaSerTyrThrLysAsnPheSerLeuGlu 20
 QY 61 AAGCTCAAGCTCAGCAGCATGAAGGATGACCTGACCGTCTCTGGCCCAATATGTGTGGC 120
 Db 21 LysLeuLysLeuSerSerMetLysAspAspLeuThrValLeuArgHisMetTrpPheGly 40
 QY 121 AGCAAGAGCGGCGATGATCAGCTGCTGCTGGCTGGAGAGCTTACGGGCCCCAGGCGCT 180
 Db 41 SerLysLysGlyAspAspHisAlaAlaArgLeuGluSerPheTyrGlyProGlnAlaAla 60
 QY 181 GCCTTT----- 186
 Db 61 AlaTyrAspAlaPheArgSerArgPheLeuTrpGlyArgArgProMetLeuAlaAlaVal 80
 QY 187 GCTGCCCGCTGGCCGAGCGCTCGAACCTCATCTGGGTTCACCTGGGTGGTGGCACTGGG 246
 Db 81 AlaAlaArgLeuAlaGluArgSerAsnLeuIleTrpValAspLeuGlyGlyThrGly 100
 QY 247 GAGAATGCTGATGATGCTGATTAATCATCGACTGGCGAAGTTCAGTCCATCTAGCTG 306
 Db 101 GluAsnValAspMetMetAlaAspTyrIleAspLeuAlaLysPheLysSerIleTyrVal 120
 QY 307 GTCGACCTGCGCACTCGCTGTGCGAGGTGCGCAAGAGAGCGCAAGGCGCAAGGGCTGG 366
 Db 121 ValAspLeuCysHisSerLeuCysGluValAlaLysLysAlaLysAlaLysGlyTrp 140
 QY 367 AAGAATGCTGAGTGTGGAGGCGCGCTGTGCAATTTGCGCCCTCGAGGCGACCGCG 426
 Db 141 LysAsnValGlnValValGluAlaAspAlaCysGlnPheAlaProProGluGlyThrAla 160
 QY 427 ACGCTCATACCTTCTCTACTGCTCATCGATGATTCACCGTTCACCAACGCTCATCGAC 486
 Db 161 ThrLeuIleThrPheSerTyrSerLeuThrMetIleProProPheHisAsnValIleAsp 180
 QY 487 CAGCTTGCTGCTGACCTGCTCCAGAGCGCTGCTGGCGTTCGCGACTTCTAGCTGAGC 546
 Db 181 GlnAlaCysSerTyrLeuSerGlnAspGlyLeuValGlyValAlaAspPheTyrValSer 200
 QY 547 GGCAAGTACGACCTGCGCCCTCGCGCAGATGCGCTGTGCGCGCGTCTTCTCGCGCATCG 606
 Db 201 GlyLysTyrAspLeuProLeuArgGlnMetProTrpSerArgPhePheTrpArgSer 220
 QY 607 ATCTTCGACATCGACACATTGACATCGGCCCGCGAGCGCGCGCTACCTGGAGCAGAAG 666
 Db 221 IlePheAspIleAspAsnIleAspIleGlyProGluArgArgAlaTyrLeuGluGlnLys 240
 QY 667 CTGGAGCGCTGTGGAGCAGAACACCCAGGGTTTCGATCCCTACGTCGCTGGCTGGC 726
 Db 241 LeuGluArgValTrpGluGlnAsnThrGlnLysSerIleProTyrValProTrpLeuArg 260
 QY 727 GCGCCCTACTACGTGTGGATTGGCGCGCTGCGCCAGCGTTCGCCACGCGCTGCACGAGGAG 786
 Db 261 AlaProTyrTyrValTrpIleGlyArgLeuProSerValGlyHisAlaLeuHisGluGlu 280
 QY 787 CGCGTGAGCGCGCGCCCATGTTCCCGCCACCTTCTGTCAGCGAGTGTGGGAGGAC 846
 Db 281 ArgValGluArgProProMetPheProProThrPheLeuTyrThrGlnSerTrpGluAsp 300
 QY 847 CCCGAGCGGATATGGAGGTGATGAGATCAACCCCAAGGACACGGTCTGACCTGCAT 906
 Db 301 ProGluProAspMetGluValMetGluIleAsnProLysAspThrValLeuThrLeuThr 320
 QY 907 AGCGCGGCTGCAATGCCCTGAACTGTGCTGGTGGAGGGCGCGCCAGGTGGTTCGGTG 966
 Db 321 SerGlyGlyCysAsnAlaLeuAsnLeuLeuValGlnGlyAlaGlyGlnValValSerVal 340
 QY 967 GACTGCAACCCCGCGAGTGGCGCTTCTGGAGCTGAAGAGTGAAGGTGGCATTCAGCAGCTG 1026
 Db 341 AspCysAsnProAlaGlnSerAlaLeuLeuGluLeuLysValAlaIleGlnGlnLeu 360

QY 1027 GAGTTTGAGACGCTGTGGCAGCTGTTTCGGCGAGGGCGTGCACCCGCGCATTTGAGGAGCTG 1086
 Db 361 GluPheGluAspValTrpGlnLeuPheGlyGluGlyValHisProArgIleGluGluLeu 380
 QY 1087 TAGCAGAAGAGCTGCGCGCTTCTCTGTCGCAAAACAGCAGCAACTTCTGGTCCAAGCGC 1146
 Db 381 TyrGluLysLysLeuAlaProPheLeuSerGlnThrSerHisAsnPheTrpSerLysArg 400
 QY 1147 CTCTGCTACTTCCAGCAGCGCTGTACTACAGGGCGCATGGCAAGCTGTGCTGGGTG 1206
 Db 401 LeuTrpTyrPheGlnHisGlyLeuTyrTyrGlnGlyGlyMetGlyLysLeuCysTrpVal 420
 QY 1207 CTCAGCTGCTGCGCTGCTGCTGGGACTGGGCAAGACCGTCAAGCGCTCGCCAAACGCG 1266
 Db 421 LeuGlnCysLeuAlaValValLeuGlyLeuGlyLysThrValLysArgLeuAlaAsnAla 440
 QY 1267 CCACAATGAGAGAGCAGCGCTCTGTGTGGGACAGCAACATGCTCATCTCATCTCGTGAAG 1326
 Db 441 ProThrMetGluGluGlnArgArgLeuTrpAspSerAsnMetLeuIleHisPheValLys 460
 QY 1327 AACGGCGCCAAAGCGCTGTGCTGCTCAAGTTCGTGAGCTGTGCTCTTCAAC 1386
 Db 461 AsnGlyProLysProLeuValTrpLeuPheValLysPheValSerLeuValLeuPheAsn 480
 QY 1387 AAGCGCTGCTGTGCTGCTGCGCGCGCTGCGCGCAAGCAGTACGCGCTGATCAAGGCG 1446
 Db 481 LysAlaValLeuTrpPheGlyGlyGlyValProGlyLysGlnTyrAlaLeuIleLysAla 500
 QY 1447 GACGGCATCCCATTTGAGAACTACATCGCGCGCGCACCATGACCGCTGGCGGAGACTCG 1506
 Db 501 AspGlyIleProIleGluAsnTyrIleAlaArgThrMetAspGlyValAlaGluAsnSer 520
 QY 1507 CAGTGGCGCAAGCAGAACTACTTCTACTACAATGCTCAACCGCTACCGCAAGTTCTCTGCGCGAC 1566
 Db 521 HisValArgLysGlnAsnTyrPheTyrTyrAsnCysLeuThrGlyLysPheLeuArgAsp 540
 QY 1567 AACTGGCCCCACCTACCTGCGCGAGCGCGCTTCGCCACCTCAAGAGTGGCGTGGTGGAC 1626
 Db 541 AsnCysProThrTyrLeuArgGluAlaAlaPheAlaThrLeuLysSerGlyValValAsp 560
 QY 1627 AACTGACCGCTCTCCAACTTCTCATGAGGAGCTCAAGGGCGCAGCTACACCAAG 1686
 Db 561 AsnLeuThrValSerThrAsnPhePheMetGluGluLeuLysAlaArgThrTyrThrLys 580
 QY 1687 GTGATTCTGATGACACACAGCTGGAGCTGCTGATATGCCCTGGCCACAGAGCTGGCGCAG 1746
 Db 581 ValIleLeuMetAspHisValAspTrpLeuAspMetProValAlaAsnGluLeuAlaGlu 600
 QY 1747 TGCTTGCCCAAGCAGGTGCGCGCGCGCATCTCATCTGCGCTCCGCTCCCTCAGC 1806
 Db 601 CysLeuAlaLysGlnValAlaProGlyGlyIleValIleTrpArgSerAlaSerLeuSer 620
 QY 1807 CCGCCTTACCGCAGCTGATCCAGAGCGCGCTTGCAGCTGGCTGCATCCGCGCGCC 1866
 Db 621 ProProTyrAlaGluLeuIleGlnLysAlaGlyPheAspValArgCysIleArgArgAla 640
 QY 1867 ACTCAGGCTTACATGACCGCTCAACATGCTACAGCTCTTCTACATGGCGCGCGCAAG 1926
 Db 641 ThrGlnGlyTyrMetAspArgValAsnMetTyrSerSerPheTyrMetAlaArgArgLys 660
 QY 1927 GGCGCCCAAGAGGACCAAC 1944
 Db 661 GlyAlaLysLysAspAsn 666

RESULT 2

Q4IPW6 GIBZE
 ID Q4IPW6 GIBZE PRELIMINARY; PRT; 835 AA.
 AC Q4IPW6;
 DT 13-SEP-2005 (TrEMBLrel. 31, Created)
 DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
 DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
 DE Hypothetical protein.
 GN ORFNames=FG00742.1;

OS Gibberella zeae PH-1.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.
 OX NCBI_TaxID=229533;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=PH-1;
 RA Birren B., Nusbaum C., Abouelleil A., Allen N., Anderson S.,
 RA Arachchi H.M., Barna N., Bastien V., Bloom T., Boguvalavkiy L.,
 RA Boukhalter B., Butler J., Calvo S.E., Camarata J., Chang J.,
 RA Choepel Y., Collymore A., Cook A., Cooke P., Corum B., DeArellano K.,
 RA Diaz J.S., Dodge S., Dooley K., Dorris L., Elkins T., Engels R.,
 RA Erickson J., Faro S., Ferreira P., Fitzgerald W., Gage D., Galagan J.,
 RA Gardyna S., Gnerre B., Graham L., Grand-Pierre N., Hafez N.,
 RA Hagopian D., Hagnon S., Hall J., Horton L., Hulme W., Iliev I.,
 RA Jaffe D., Johnson R., Jones C., Kamal K., Kato A., Karatas A.,
 RA Kells C., Landers T., Levine R., Lindblad-Toh K., Liu G., Lui A.,
 RA Ma L.-J., Mabbitt R., Maclean C., Macdonald P., Major J., Manning J.,
 RA Matthews C., Mauceli E., McCarthy M., Meldrim J., Meneus L.,
 RA Minova T., Mlenga V., Murphy T., Naylor J., Nguyen C., Nicol R.,
 RA Nielsen C.B., Norbu C., O'Connor T., O'Donnell P., O'Neill D.,
 RA Oliver J., Peterson K., Phunkhang P., Pierre N., Purcell S.,
 RA Rachupka A., Ramasamy U., Raymond C., Retta R., Rise C., Rogov P.,
 RA Roman J., Schauer S., Schupbach R., Seaman S., Severy P., Smirnov S.,
 RA Smith C., Spencer B., Stange-Thomann N., Stojanovic N., Stubbs M.,
 RA Talamas J., Testaye S., Theodore J., Topham K., Travers M.,
 RA Vassiliev H., Venkataraman V.S., Viel R., Vo A., Wang S., Wilson B.,
 RA Wu X., Wyman D., Young G., Zainoun J., Zembek L., Zimmer A., Zody M.,
 RA Lander E.,
 RT "Fusarium graminearum genome sequence."
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
 CC -!- CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 DR EMBL: AACM0100035; EAA70688.1; -: Genomic_DNA.
 KW Hypochemical protein.
 SQ SEQUENCE 835 AA; 93926 MW; 819222A79ABC086E CRC64;

Alignment Scores:
 Pred. No.: 4,84e-41 Length: 835
 Score: 953.00 Matches: 254
 Percent Similarity: 45.9% Conservative: 103
 Best Local Similarity: 32.7% Mismatches: 216
 Query Match: 26.3% Indels: 204
 DB: 2 Gaps: 25

US-10-620-914-44 (1-1947) x Q4IPW6_GIBZE (1-835)

QY 118 GGCGAAGAGGGGATGATCATCGCTGCTGGAGAGCTTCTACGGGCCCGCCAGGCC 177
 DB |||||
 64 GlyAspSerLysGlyAsnGlnAspAla---LeuGluSerPheTyrLysLysGlnAla 82
 |||||
 QY 178 GCTGCTTT----- 186
 |||||
 DB 83 GlyAlaTyrAspAlaThrArgLysValLeuLeuArgGlyArgGluAspMetLeuAlaLeu 102
 QY 187 ---GCTGCCCGCTGGCC-----GAGCGCTCGAACCTCATCTGGGTT 225
 |||||
 DB 103 ValAlaAlaGlnMetGlnAlaLysLeuLysAspAsnLysLysThrLysArgValTrpVal 122
 QY 226 GACCTGGGTGGTGGCACTGGGAGAATGTCATATGATGCTGCTGATTCATCGACCTGGCG 285
 |||||
 DB 123 AspIleGlyGlyThrGlyPheAsnIleGluAlaMetGlyAlaPheValaspValPro 142
 QY 286 AAG---TTCAAGTCCATCTACGTGCTGACCTGCTGCCATCTGCTGCGAGGTGGCAAG 342
 |||||
 DB 143 ThrPhePheSerValTyrLeuValaspPheSerProSerLeuCysGluValAlaLys 162
 QY 343 AAGAAAGCGAAGCGCAAGGCTGGAAGAATGTCAGGTCTGTGGAGCGCAGCGCTTGCAC 402
 |||||
 DB 163 LysArgPheAspArgLeuGlyTrpLysAsnValLysValValCysGluaspAlaArgLys 182
 |||||
 QY 403 TTT-----GCGCCCGCTGAG----- 417

183 PheArgLeuGluAspTyrGluSerGlyMetProSerLysSerIleProProArgSerPro 202
 418 -----GGCACCGCGACGCTCATCACCTTC 441
 203 AlaLeuSerTyrPheAspLysProArgProAspPheGlyAlaAspLeuIleThrMet 222
 442 TCCTACTCGCTCACGATGATTCACCGTTCCACACGTCATCCACGAGGCTTCCTCTGCTAC 501
 223 SerTyrSerLeuSerMetIleProAspTyrTyrSerValIleAspSerValThrSerLeu 242
 502 CTGTCCCAAGACGCTGTGGGCGCTTCCGACTTCTACGTGAGCGCAGTACGACCTG 561
 243 LeuSerProGlnGlyIleMetGlyValValaspPheTyrValGlnAsnLysValaspPhe 262
 562 CCCTG-----CGCCAGATGCCCTGGTGGCGCGCTTC 594
 263 AlaPheArgAsnTyrThrGlyGlyLeuValaspArgHisValAsnPheLeuSerArgSer 282
 595 TTCTGGCGATCGATCTTCGACATCGACATCGCCCGCGAGCGCCGCGCTAC 654
 283 PheTrpArgSerTrpPheAspLeuAspArgValGlyLeuGluProSerArgArgAspTyr 302
 655 CTGGAGCAGAGCTGGAGCGCTGTGGAGCAGACACCCAG-----GGTTCG 702
 303 LeuGluTyrLysPheGlyThrValLeuAsnValAsnThrArgAsnLysGlyLeuGlyAla 322
 703 ATCCCTACGTGCGGTGGCTGCGCGCCCTCTACTGTGGATTGCG-----CGCTG 756
 323 Ile-----ProTyrTrpIleTrpLeuGlyCysHisLysLys 334
 757 CCCAGCTTGGCCACCGCTCGACGAGGAGCGGTGGAGCGG----- 798
 335 ProPheSerSerSerSerLeuProHisGluValGluArgIleAspAlaLeuValThr 354
 799 -----CCGCCCATGTTCCCG----- 813
 355 GluSerProTyrLeuTyrProAlaAsnHisGlyAspAlaLeuThrArgAlaIleGluArg 374
 813 ----- 813
 375 SerAlaProGluIleArgSerLysAlaPheLeuThrAlaValSerAsnLeuSerSerAsn 394
 814 -----CCACCTTCCTGTAC----- 828
 395 LeuProLeuProSerPhePheTyrGlnAsnHisTrpArgIleTyrTyrAspGluGln 414
 829 -----ACGCAG-----TCGTGGAGGACCCC 849
 415 LeuProLysHisThrGlnPheLysAspGluTyrIleTyrAlaPheThrTrpGluAspThr 434
 850 GAGCCGATATGGAGTGGATGGAGATCAACCCCAAGCAGACAGCTGTGACCTGACTAGC 909
 435 ArgValaspGluArgIleLeuLysLeuGlyAlaAspAspLysValLeuAlaIleThrSer 454
 910 GGCGGTGCAATGCCCTGAACCTGTGTCGAGGGCGCGCCGAGGTGTGTGCTGGTGGAC 969
 455 AlaGlyAspAsnIleLeuSerTyrLeuLeuGlnSerProAlaArgValHisAlaValasp 474
 970 TGCAACCCCGCGAGTCCGGCGCTTCTGGAGCTGGAAGAAGTGGCCATTTCAGCAGCTGGAG 1029
 475 LeuAsnProThrGlnAsnHisLeuLeuGluLeuLysAlaAspSerTyrThrAlaLeuPro 494
 1030 TTGTGAGAGTGTGGCAGCTGTTCGGCGAGGGGTGCACCCCGCATTTGAGGAGCTGTAC 1089
 495 TyrGlyAspPheTrpLysIlePheGlyAspGlyLysHisProArgPheArgGluLeuLeu 514
 1090 GAGAAAGAGCTGGCGCTTCCTCTGTCGCAACACCGACCAACTTCTGCTCAAGCGCTC 1149
 515 IleThrLysLeuSerProHisLeuSerGlyArgAlaPheGlnTyrTrpLeuLysAsnVal 534
 1150 TGGTACTTTCAG-----CACGGCTGTACTACAGGGCGGCATGGGCAAGCTG 1197
 |||||

Db 239 GlyGlyThrGlyTrpAsnIleGluGlnMetAsnGlnPhePheProIleAspGlnLeuSer 258
QY 295 TCATCTACGTGTCGACCTGTCCTGTCGTCGAGGTGGCAGAGAGAGCGCAAG 354
Db 259 GlnValTyrLeuIleAspLeuGluProLeuLeuGlnValAlaArgLysArgPheAla 278
QY 355 GCCAAGGCGTGGAGATGTCAGGTGTCGAGCGCCGACCTTGCCTGATTTGGCCCT 414
Db 279 AlaLysGlyPheLysAsnValGlnValLeuGluCysGlnAspAlaSerGlnPheAsnMetPro 298
QY 415 GAGGCGACCGCGCAGC-----CTCATCACCTTCTCTCTCTCTCTCTCTCTCT 459
Db 299 ---GlyLeuAlaAlaGlyGlnLysValAspLeuPheThrCysSerTyrSerIleSerMet 317
QY 460 ATTCCACCGTCCACACGATCATGACAGCTGTCGTCGACCTGTCCTCCAA---GACGGC 516
Db 318 IleProPheTyrAlaValLeuAspArgIleAsnAspLeuLeuAspProValThrGly 337
QY 517 CTGGTGGCGTTCGCGACTTCTAGCTGAGCGGCAAGTACACCTGCCCTG----- 567
Db 338 ValPheGlyValValAspPheTyrValSerGlySerGly---ProLeuAlaLysSer 356
QY 568 -----CGCCAGATGCCCTGCTGCTGCGCGCTTCTCTCTGCGCA 603
Db 357 ProMetIleGlyAspThrArgArgGlnCysGlyTyrLeuSerArgIlePheTrpSer 376
QY 604 TCATCTTCGACATCGACATTCATGATCGCGCGCGCGCGCTACCTGGAGCAG 663
Db 377 MetTrpPheSerPheAspHisIleGluLeuHisProAlaArgAspTyrLeuGluHis 396
QY 664 AAGCTGGAGCGGTGGGAGCAGACACCCAGGTTGCG-----ATCCCTACGTGCGG 717
Db 397 LysPheGlyThrIleLysCysTyrAsnGlyArgAsnAsnPheIleIleProPheIle--- 415
QY 718 TGCTGCGCGCCCTACTACTGCTGATGGC-----CGCTGCGCGCGCTGGCCAC 771
Db 416 ---ValArgIleProTyrIleIleTrpLeuGlyValSerArgGluArgPheThrLys 434
QY 772 GCCTGTCAGGAGCGCGTGGAGCGG-----CGCCCATGTTCCCG 813
Db 435 AlaIleGlnAlaPheGluValGluSerGlyAsnArgValValProSerPhePro 454
QY 814 CCC----- 816
Db 455 GluLeuAlaTyrMetHisAsnGlyGluAlaHisThrHisSerIleSerAlaAlaThr 474
QY 816 ----- 816
Db 475 AlaThrAlaThrAlaThrGlyIleAspAlaAspGluAspAlaSerSerArgPhe 494
QY 816 ----- 816
Db 495 GluProAlaArgArgLeuLeuArgArgValSerGluAlaThrThrGluSerAspAla 514
QY 816 ----- 816
Db 515 AspSerAspArgProLeuLysLeuGluLeuGlyProHisPheProLeuSerSerPheHis 534
QY 816 ----- 816
Db 535 TyrGlnLysArgGlnTrpArgLeuProPheValAspAsnGluPheSerAspMetPheArg 554
QY 817 ACCTTCTGTACAGCAGTGTGGGAGGACCCGAGCGGATGATGGAGTGTGGAGATC 876
Db 555 ThrTrpIleTyrGlyPheThrTrpGluAspProTyrValAspMetGlnHisLeuAspLeu 574
QY 877 AACCCCAAGACACGGTGTGACCTTCAGTACGCGCGGTGCAATGCTTCAACCTGCTG 936
Db 575 GlyLysAspAspSerIleLeuCysIleThrSerAlaGlyAspAsnAlaLeuHisTyrAla 594
QY 937 GTCACAGGG---CGCCGCGCAGGTGTGCTGCTGCACTGCAACCCCGCGAGTGGCGCT 993
Db 595 ValAlaGlyLysProArgArgIleHisAlaValAspMetAsnProCysGlnGlyHisLeu 614

RESULT 4

Q6C6E3_YARLI

ID Q6C6E3_YARLI PRELIMINARY; PRT; 741 AA.

AC Q6C6E3;

QY 994 CTGAGCTGAAGAAGTGGCCATTCAGCAGCTGGAGTTTGAGGACGTGTGCGAGCTGTTTC 1053
Db 615 LeuGluLeuLysLeuAlaCysIleAlaSerLeuSerTyrAspGluMetTrpGlnMetPhe 634
QY 1054 GCGAGGCGGTGCACCGCGCATTTAGGAGCTGTACGAGAGAGCTGGCGCGCTTCTCTG 1113
Db 635 GlyGluGlyArgIleAspAsnPheArgGluLeuAspSerLysIleSerProTyrLeu 654
QY 1114 TCCAAACAGCAGCCACCACTTCTGCTCCAAGCGCTCTGTGTACTTTCAGACGCGCTGTAC 1173
Db 655 SerSerHisAlaTyrGlnPheTrpArgLeuAsnThrArgAlaPheAspLysAlaPheTyr 674
QY 1174 TACAGGCGGCGATGGGCAAGCTGTGCTGGTGTCTGAGTGTGCTGCGCTGTCTGGGA 1233
Db 675 PheArgGlyTyrSerGlyHisAlaLeuArgLeuAlaLysPheAlaPheSerValThrGly 694
QY 1234 CTGGGCAAGACCGTCAAGCGCTCCGACCGCCGACCAATGAGGAGGAGCGCGCTGTG 1293
Db 695 ValArgArgTrpValGluLysMetCysThrAlaAsnSerValGluGluGlnGlnVal 714
QY 1294 TGGGACAGCAACATGCTCATCTCCACTTCTGTCGAAGACGGCGCCCAAGCGCTGGTGTGCTG 1353
Db 715 TrpAspLysLysLeuArgSerThrLeuIleAsn-----LysProLeuIleArgLeu 731
QY 1354 TTCGTCAAGTTCGTGAGCCTGTGCTTCAACAGCGCGCTGTGCTGTGGTGTGGCGCGC 1413
Db 732 Phe-----LeuSerAsnProAlaPheLeuTrpAsnAlaLeuGly 744
QY 1414 GTGCGCGGCAAGCAGTACCGCTGTATCAAGCGGAGCGCATCCCATTTGAGAACTACATC 1473
Db 745 ValProMetAsnGlnTyrGlnIlePheLeuAsnGluGlyValSerAlaGluGlnPheAla 764
QY 1474 GCGCGCACATGACGCGCTGCGGAGAACTCCGACGTGCGCAAGCAGAACTACTTCTAC 1533
Db 765 IleAspThrLeuAspSerIleProSerArgSerLeuLeuLysAsnAspAsnTyrHisTyr 784
QY 1534 TACAACCTGCTCACCGGCAAGTTCCTGCGCGACAACTGCGCCACCTACCTGCGCGAGCG 1593
Db 785 GlnLeuCysLeuLeuHisLysTyrThrLysGlnSerCysProLeuTyrLeuLysProAsp 804
QY 1594 GCCTTCGCCACCTCAAGAGTGGCGTG-----GTGACCAACTGACCGTCTCC 1641
Db 805 GlyPheAlaAlaLeuLysLysGlnAlaLeuGlnAspGlyLeuAspSerPheArgLeuHis 824
QY 1642 ACCAATCTTCTCATGAGGAGCTCAAGCG-----CGCACCTACACCAAGGTATT 1692
Db 825 ThrAspSerIleValAsnValLeuArgGlyPheGluAspGlyAlaLeuThrArgAlaIle 844
QY 1693 CTGATGACACCGCTGCACTGGCTGGAT---ATGCCCGTGGCC----- 1731
Db 845 ThrMetAspHisMetAspTrpPheAspProValProAlaSerArgProAlaProThrIle 864
QY 1732 -----AACGAGCTGGCGAG 1746
Db 865 LysGlnAlaArgAspSerAspLysSerValSerAspLeuAspArgGluIleCysGlu 884
QY 1747 TGCTGCGCAAGCAGGCTGCGCGCGCGCATCGTATCTGCGCGCTCCGCGCTCCCTCAGC 1806
Db 885 ---LeuSerArgValIleArgLysGlyAlaValPheTyrArgSerAlaAlaLysLys 903
QY 1807 CGCCCTACCGCAGCTGATCCAGAGCGGCGCTTCGAGTGGCTGCAATCCGC---CGC 1863
Db 904 ProTrpTyrAsnHisArgPheGluLysMetGlyPheSerValGlnProValHisIleArg 923
QY 1864 GCCACTCAGGCTACATGACCGCGCTCAACATGTACAGTCTCTTACATGGCGCGCG 1923
Db 924 GluThrAlaLysProIleAspAsnValAsnMetTyrAlaSerPheTyrLysAlaThrArg 943
RESULT 4
Q6C6E3_YARLI
ID Q6C6E3_YARLI PRELIMINARY; PRT; 741 AA.
AC Q6C6E3;

QY 819 ----- 819
 Db 322 ThrProIleAlaAsnGlnLeuGluAspIleProIleSerLysGlyHisGluAlaAlaLeu 341
 QY 819 ----- 819
 Db 342 IleAsnLeuGlnLysAsnLeuProTyrProSerMetTyrTyrGlnLysGluTyrTrpArg 361
 QY 820 ----- 820
 Db 362 ValTyrTyrAspGluMetAsnProLeuTyrGluGlnPheLysAsnGlnTyrIleTyrAla 381
 QY 832 CAGTCGTGGAGGAGCAGCCGAGCGGATATGAGGTGATGAGATCAACCCCAAGGACAGC 891
 Db 382 PheThrTrpGluAspProArgGluAspHisLysLeuAsnPheThrSerAspAspThr 401
 QY 892 GTGCTGACCTGACTAGCGCGGCTGCAATGCCCTGAACCTG-----CTGGTGCAGGG 945
 Db 402 ValLeuAlaIleThrSerAlaGlyAspAsnIleLeuSerTyrAlaSerLeuProThrPro 421
 QY 946 GCCGGCCAGTGTGTGCTGGTGGACTGCAACCCCGCAGTCGCGGCTTCTGGAGCTGAAG 1005
 Db 422 ProLysLysIleHisAlaValAspLeuAsnProCysGlnAsnHisLeuGluLeuLys 441
 QY 1006 AAGTGGCCATTGACGAGCTGGAGTTTGGAGCAGCTGTGGCAGCTGTTCGGCGAGGGCGTG 1065
 Db 442 LeuAlaSerPheArgCysLeuSerGlnGluGlnIleTrpSerMetPheGlyGluGlyLys 461
 QY 1066 CACCCGCGCATTGAGGAGCTGTACGAGAAGAGCTGGCGCCCTTCTGTCGCAACACGAG 1125
 Db 462 IleGluAsnPheAsnAspLeuLeuIleAspThrLeuAlaProHisMetSerSerAsnAla 481
 QY 1126 CACACTTCTGCTCCAGCGC-----CTCTGTGTACTTCCAGCAGCGCGCTGTACTACAG 1179
 Db 482 PheGlnTyrTrpMetAspLysGlyProLysThrPheSerGlyLysGlyLeuTyr----- 499
 QY 1180 GCGGCATGCGAAGCTGTGCTGGTGTCTGCGAGTGC-----CTGGCGGTGGTCTG 1230
 Db 500 ---AspThrGlyPheSerArgTrpAlaLeuArgLeuSerArgTyrValPheLysValCys 518
 QY 1231 GGACTGGGCAAGCCGTCAAGCGCTGCGCCACCGCCCAATGGAGGACGACGCGCT 1290
 Db 519 GlyValSerLysTyrValGluGluLeuCysAlaAlaThrThrMetGluGluGlnLeuArg 538
 QY 1291 CTGTGGACACCAATCTCTCATCTCCATCTGCTGAAGACGGCGCCAGCGCTGTGTGG 1350
 Db 539 IleTrpAsnGluHisLeu-----LysProThr----- 547
 QY 1351 CTGTGTCTCAAGTTCTGTG---AGCTGTGTCTTCAACAGCGCGTGTGTGTCTGCGC 1407
 Db 548 LeuPheAsnProValValGlySerLeuLeuValGlyAsnProMetPheLeuTrpLysAla 567
 QY 1408 GCGCGGTGCGGCAAGCAGTACGCGCTGATCAAGCGGAGCGCATCCCATTTGAGAAC 1467
 Db 568 LeuGlyValProAlaAsnGlnAlaLeuMet-----GlyProSerValIleLys 584
 QY 1468 TACATCGCGCCACATGAGCGGTGCGGAGAACTGCGACGTGCGGACGACCAATAC 1527
 Db 585 TyrValValAspThrLeuAspProIleIleLysArgSerMetIleSerAsnAspAsnTyr 604
 QY 1528 TTCTACTACAACTGCCTCAGCGGCAAGTTCTGCGCGCACTACCCACCTACTGCGC 1587
 Db 605 PheTyrTyrLeuCysMetMetGlyArgTyrThrLysAsnAsnCysProAspTyrLeuThr 624
 QY 1588 GAGCGCGCTTGGCCACCCCTCAAG-----AGTGGCGTGGTG 1623
 Db 625 ThrLysGlyPheAsnArgLeuSerSerThrAlaAlaThrAlaSerGlySerSerProIle 644
 QY 1624 GACAACTGACCGTCTCCACCAACTTCTTCATGAG-----GAGCTCAAGCGCGC 1674
 Db 645 AspAsnLeuArgIleHisThrAspThrLeuAsnGluValPheGlyArgLeuLysGluLys 664
 QY 1675 ACCTACCAAGGTGATTTCTGATGACCAACCGTGGACTGGCTGGAT-----ATGCC 1725

Db 665 SerIleThrIleAlaIleMetAspHisMetAspTrpPheAspProAsnGlyArgAsp 684
 QY 1726 GTGGCCAAAGAGTGGCCGAGTCCCTGCGCAAGCAGGTTGCGCGCGGCATCGTCATC 1785
 Db 685 AlaIleAsnGluIleThr---AlaLeuLysArgCysLeuAlaProGlyGlyArgValLeu 703
 QY 1786 TGCGCGTCCGCGCTCCCTCAGCCCGCTACGCCGAGCTGATCCAGAGCGCGGCTTC--- 1842
 Db 704 LeuArgSerAlaSerThrLysProTyrTyrLeuLysThrPheLysAsnLeuGlyPheGln 723
 QY 1843 GAGCTGCGTGTGATCCGCGCGCCATCTCAGGGGTACATGACCGCGTCAACATGTACAGC 1902
 Db 724 GluGluGluAsnValValArgGlnProGlySerSerIleAspArgValAsnMetTyrAla 743
 QY 1903 TCC 1905
 Db 744 Asn 744
 RESULT 6
 Q6CJ18_KLJ18
 ID Q6CJ18_KLJ18 PRELIMINARY; PRT; 790 AA.
 AC Q6CJ18; 25-OCT-2004 (TREMBlrel. 28, Created)
 DT 25-OCT-2004 (TREMBlrel. 28, Last sequence update)
 DT 25-OCT-2004 (TREMBlrel. 28, Last annotation update)
 DE Similar to ca|CAL783|pf7635 Candida albicans unknown function.
 GN OrderedLocusNames=KLJ180F22198g;
 OS Kluyveromyces lactis (yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
 OX NCBI_TaxID=28985;
 RN [1]
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA]
 RC STRAIN=CBS 2359 / IFO 1267 / NRRL Y-1140 / WM37;
 RX PubMed=15229592; DOI=10.1038/nature02579;
 RA Dujon B., Sherman D., Fischer G., Durand P., Casaregola S.,
 RA Lafontaine I., de Montigny J., Marck C., Neuvéglise C., Talla E.,
 RA Goffard N., Franchin S., Beckerich J.-M., Beyne E., Bleykasten C.,
 RA Barney S., Blanchin S., Beckerich J.-M., Confanier F., de Daruvar A.,
 RA Boissrame A., Boyer J., Fairhead C., Ferry-Dumazet H., Groppi A.,
 RA Despons L., Fabre E., Fairhead C., Jaumaux N., Joyet P., Kachouri R.,
 RA Hantraye F., Hennequin C., Jaumaux N., Lesur I., Ma L., Muller H.,
 RA Kerret A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
 RA Nicoud J.-M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
 RA Pellenz S., Potier S., Richard G.-F., Straub M.-L., Suleau A.,
 RA Swennen D., Tekala F., Wesolowski-Jouvet M., Westhof E., Wirth B.,
 RA Zeniou-Meyer M., Zivanovic Y., Bolotin-Fukuhara M., Thierry A.,
 RA Bouchier C., Caudron B., Scarpelli C., Gallardin C., Weissenbach J.,
 RA Wincker P., Souciet J.-L.;
 RT "Genome evolution in yeasts";
 RL Nature 430:35-44(2004).
 DR EMBL; CR382126; CAG98779.1; -; Genomic DNA.
 DR GO; GO:0008757; F:S-adenosylmethionine-dependent methyltransf. .; IEA.
 DR GO; GO:0016740; F:transferase activity; IEA.
 DR InterPro; IPR000051; SAM bind.
 KW Complete proteome; Methyltransferase; Transferase.
 SQ SEQUENCE 790 AA; 90212 MW; 25BA985E9C5B0D1 CRC64;
 Alignment Scores:
 Pred. No.: 8,14e-38 Length: 790
 Score: 890.50 Matches: 233
 Percent Similarity: 46.6% Conservative: 94
 Best Local Similarity: 33.2% Mismatches: 241
 Query Match: 24.6% Indels: 133
 DB: 2 Gaps: 23
 US-10-620-914-44 (1-1947) x Q6CJ18_KLJ18 (1-790)
 QY 151 CTGAGAGCTTCTAGCGGCGCCCGCGCTGCTTT----- 186
 Db 93 LeuGluGlnPheTyrLysSerGlnAlaLysLeuTyrAspArgThrArgGlyValLeuLeu 112

187 QY -----GCTCCCGCTGGCGGAGCGCTGCAACCTC 216
113 Db GlnGlyArgGluThrSerLeuLysLeuSerLeuSerHisLeuSerGluLysLysGlyAsn 132
217 QY ATCTGGGTGACCTGGGTGGGACCTGCGGAGAAATGTCATATGATGGCTGATTACATC 276
133 Db ValTrpIleAspValGlyGlyThrGlyPheAsnIleSerGlnMetAlaLeuLeuThr 152
277 QY GACCTG---GCGAAGTTCAAGTCCATCTACGTGGTCCAGCTGTGCCACTCGCTGTGCGAG 333
153 Db AsnLeuAspThrThrPheAspLysIleTyrLeuIleAspLeuSerProSerLeuCysGlu 172
334 QY GTGGCCAGAGAGCGAGGCGAGCGGCTGGAGAGATGTCAGGTCGTGGAGCGCGAC 393
173 Db ValAlaArgLysCysGlySerGluHisGlyTrpLysAsnValGluValIleCysGlyAsp 192
394 QY GCTTGCCAAATTGGCCCTCGAGGCGACCGCGAGCTCATCCTTCTCTACTCTGCTC 453
193 Db AlaCysAspPheGluIleProGluLysSerAlaGlnLeuIleThrPheSerTyrSerLeu 212
454 QY ACGATGATTCACCGTTCCACAACGTATCGACAGCGCTTGCTGTACCTG---TCCCAA 510
213 Db SerMetIleProSerPheTyrAlaAlaIleAspHisAlaValSerLeuLeuAspAlaLys 232
511 QY GACGGCTGTGGCGCTGGCGACTTCTACGTGAGCGCGAAGTAC----- 555
233 Db AsnGlyIleIleSerCysValAspPheGlyValThrAsnGluSerMetLeuValGlyArg 252
556 QY -----GACCTGCCCTCGCGCAGATGCCCTGGTCGCGCGCTTCTCTG 600
253 Db ThrAsnThrLeuGlyGlyLeuValAsnArgHisIleProTrpLeuPheArgThrPheTrp 272
601 QY CGATCGATCTTCGACATCGACAACATGACATCGGCGCGCGCGCGCTACCTGGAG 660
273 Db ArgLeuTrpPheGluPheAspLysValPheLeuAspProAlaArgGluTyrLeuGlu 292
661 QY CAGAGCTGAGCGGTGGGAGCAGAACACCCAG-----GGTTCGATCCCC 708
293 Db TyrArgPheGlyThrIleLysSerLeuAsnCysTyrAsnTyrLysLeuGlyLysIlePro 312
709 QY TACGTGCGGTGGGTG----- 723
313 Db TyrTrpIleTrpLeuGlyCysAsnLysAspHisGluGlnHisLeuGlnAlaArgPheVal 332
724 QY -----CGCGCCCTACTACGTGGATTGGCGC----- 753
333 Db GluLeuAlaProThrSerProTyrLeuAlaProIleThrThrSerAlaSerSerAsnAla 352
754 QY -----CTG 756
353 Db GlnProMetThrLysAlaMetIleAlaLeuGluAsnSerLysLysGlyLeuProTyr 372
757 QY CCACGCTTGGCCAGCGCTGACGAGAGCGGTG-----GAGCGCGCGCC--- 804
373 Db ProSerLeuPheTyrGlnLysGluHisTrpArgValTyrTrpAspGluValAsnProGlu 392
805 QY -----ATGTTCCCGCCACTTCTGTATACCGAGTCGTGGGAGGACCCCGCGCGAT 858
393 Db TyrSerGlnPheLysAsnSerTyrIleTyrAlaPheThrTrpGluAspProArgGluAsp 412
859 QY ATGAGGTGATGAGATCAACCCAGGACGCGTGTGACCTGACCTAGCGCGCGGTGC 918
413 Db ValAsnIleLeuAsnIleGlnProGluAspThrIleLeuAlaIleThrSerAlaGlyAsp 432
919 QY AATGCTCTGAAC-----CTGCTGGTGTGAGGGCGCGCGAGGTGTGCTGGTGGAGTGC 972
433 Db AsnIleLeuHisTyrAlaThrLeuProAsnProProLysArgIleHisGlyValAspLeu 452
973 QY AACCCCGCGCAGCTGGCGCTTCTGGAGCTCAAGAAGGTGGCCATTTCAGCAGCTGGAGTTT 1032
453 Db AsnProCysGlnGlyHisLeuThrGluLeuLysLeuAlaIleArgSerLeuSerPhe 472
1033 QY GAGGACGTGTGGCAGCTGTTCGGGAGGGCGTGCACCCCGCGCATTTGAGGAGCTGTACGAG 1092

RESULT 7

Q52GF6 MAGGR
ID Q52GF6 MAGGR PRELIMINARY; PRT; 832 AA.
AC Q52GF6
DT 13-SEP-2005 (TRENBLrel. 31, Created)
DT 13-SEP-2005 (TRENBLrel. 31, Last sequence update)

473 Db ThrGlnLeuTrpGlnMetPheGlyGluGlyLysIleAspArgPheAsnAsnIleLeuLeu 492
1093 QY AAGAAGCTGGCGCTTCTCTGTCGCAAAACCCAGCACCACTTCTGGTCCAAAGCGCTCTGG 1152
493 Db AsnLysLeuAlaProTyrLeuSerSerAsnAlaPheGlnTyrTrp----- 507
1153 QY TACTTCCAGCAGCGCTG-----TACTACAGGGCGGC 1185
508 Db ---PheGluAsnGlyThrLysThrPheAspProAsnGlyAlaGlyLeuTyrAspThrGly 526
1186 QY ATGGGCAGCTGTGCTGGGTGCTGCAGTGC-----CTGGCCGTGGTCTGGGACTG 1236
527 Db PheThrLys-----TrpAlaLeuArgLeuAlaLysTrpValPheLysValAlaAsnLeu 544
1237 QY GGCAAGACCGTCAAGCGCTCGCAACCGCCCAACATAGAGAGAGAGCGCGCTCTCTGG 1296
545 Db ThrAspGluValAsnMetLeuCysLysAlaLysThrLeuGluGluGlnArgSerIleTrp 564
1297 QY GACAGCAACATGCTCATCTCCACTTCTGTAAGAACCGGCGCCCAAGCGCTGTGTGCTGTC 1356
565 Db AspLysLysIle-----LysProVal-----LeuPhe 573
1357 QY GTCAAGTTCGTGAGCTGTGTGCTCTTC---AACAGCGCGCTGTGTGTGCTGGCGGCGC 1413
574 Db AsnArgValValGlyLysIleLeuValGlyAsnProLeuPheLeuTrpSerAlaLeuGly 593
1414 QY GTCCCGGCGAAGCAGTACGCGCTGATCAAGCGGAGCGCATCCCATTTGAGAACTACATC 1473
594 Db ValProArgAsnGln-----AlaLysMetMetGlySerSerThrLeuGlnTyrIle 610
1474 QY GCGCGCACCATGACCGCGGTGCGGAGAACTCGACGTGCGGAGCGAAGCAACTACTTCTAC 1533
611 Db IleAspThrLeuAspProValIleAspAsnSerLeuIleSerAspAspAsnTyrPheTyr 630
1534 QY TACAACTGCTCACCAGTTCCTGCGGACAACTGCCACCTACCTGCGCGGCGGCG 1593
631 Db TyrLeuThrLeuLysGlyArgTyrSerSerArgSerCysProAspTyrLeuLysGluGly 650
1594 QY GCCTTCGCGCACCTCAAG-----AGTGGGTGTGTGACCAACCTGACCGCTCTCC 1641
651 Db GlyPheLysSerLeuSerArgGluSerProGluSerProLeuAspArgValArgLeuHis 670
1642 QY ACCAACTTCTTC-----ATGAGAGAGCTCAAGCGCGCACCTACACCAAGGTGATT 1692
671 Db ThrAspThrLeuLysAspValCysGluArgLysSerLysLysThrValSerIleAlaIle 690
1693 QY CTGATGACACCACTGCACTGGCTGGATATGCCGTGCGCCACAGCTGGCGGAG----- 1746
691 Db IleMetAspHisMetAspTrpPheAsp---ProGlnGlyThrAspValAspGluIle 709
1747 QY ---TGCCTGCCAAGCAGCGTTCGCGCGCGCATGCTCATCTGCGCTCGCGCTCTCCCTC 1803
710 Db GlnAlaLeuTrpLeuAlaLeuAsnSerArgGlyArgValLeuLeuArgSerAlaSerLys 729
1804 QY AGCCCGCTTACCGCGAGCTGATCCAGAGCGGCGGTTCGACGTGCGCTGCATCCGCGCG 1863
730 Db SerProTrpTyrIleLysAsnPheGluLysPheSerCysLysAlaValSerAla 749
1864 QY GCCACTCAGGCG---TACATGGACCGCGTCAACATGTACAGCTCTTCTACATGGCGCGC 1920
750 Db ArgTyrProGlyLysCysIleAspArgValAsnMetTyrAlaSerThrTrpValCysGln 769
1921 QY CGG 1923
770 Db Lys 770


```
QY 898 ACCGTGACTAGCGCGCTGCAATGCCCTGAACTGCTGGTGCAGAGGGCGCGCCAGGTG 957
Db 399 AlaileThrSerAlaGluAspAsnLeuSerTyrlleSerGlnSerProAlaVal 418
QY 958 GTGTCGTGACTGCAACCGCGCAGTCGCGCTTCTGGAGCTGAAGAAGGTGGCCATT 1017
Db 419 HisAlaValAspLeuAsnProSerGlnAsnHisLeuLeuGluLeuLysLeuAlaSerTy 438
QY 1018 CAGCAGCTGAGTTGAGGACGTGTGGCAGCTGTTCGGCAGGCGGTGCACCGCGCATT 1077
Db 439 SerAlaLeuAspTyrlleAlaAspPheTrpLysilePheGlyGluGlyHisAspGluPhe 458
QY 1078 GAGGAGCTGACGAGAGAGTGGCGCCCTTCTGTCGCAACACGACCACTTCTGG 1137
Db 459 ArgAlaLeuLeuLeuAsnLeuSerProHisLeuSerSerArgAlaPheGlnTrp 478
QY 1138 -----TCCAAGCGCTCTGTACTTCCAGCAGCGCTGTACTACACGCGCGCATGGC 1191
Db 479 PheAlaAsnAlaHisilePheThrSerLysHisGlyLeuTrpAspThrGlyGlySerLys 498
QY 1192 AAGCTGTGCTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1251
Db 499 HisAlaileAlaAlaPheArgTrpIleSerSerAlaPheGlyValArgSerAlaValArg 518
QY 1252 CGCTCGCCCAACGCGCCCAATGAGGAGGAGCGCGCTGCTGCGAGCAGCAACATGCTC 1311
Db 519 GluMetLeuGluAlaLysThrLeuAsnGluGlnArgGluileTrpGlnGlyArgile-- 537
QY 1312 ATCCACTTCTGAGAGACGGCGCCCAACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1371
Db 538 -----ArgProAla-----LeuLeuSerArgMetValSer 547
QY 1372 ---CTGTGTGCTTCAACAGCGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1428
Db 548 AsnPheValValSerGlnGluGlnPheLeuTrpSerAlaLeuGlyValProLysHisGln 567
QY 1429 TAGCGCTGTATCAAGCGCGAGCGGATC----- 1455
Db 568 LeuAlaMetLeuGluArgAspHisAlaThrSerLysValValSerArgGlyGluAlaSer 587
QY 1456 -----CCATTGAGAACTACATCGCGCGCACCATGACGCGCGTGGCGGAG 1500
Db 588 AsnSerArgThrSerAlaileTrpGlnPheMetSerAspThrLeuAspProValAlaHis 607
QY 1501 AACTGCGCAGTGGCGAGAGAACTACTTCTACTACAACTGCTCTACCGCGCAAGTTCCTG 1560
Db 608 AspThrHisileAlaGluGluAsnProTyrlleValThrMetAlaGlyLysPheSer 627
QY 1561 CGGACAACTGCGCCCACTACCTGCGGAGCGCGCTTCCGCCACCTC-----AAGAGT 1614
Db 628 ProArgCysHisProAspTyrlleLeuSerProAlaAlaHisAlaLysLeuSerArgGly 647
QY 1615 GCGTGTGGACAACTGACCGCTCCACC-----AACTTCTTCATGGAGGAGCTC 1665
Db 648 AlaLeuGluAspArgileArgileHisThrGluLeuLeuAspValValLeuAlaGlyLeu 667
QY 1666 AAAGCGCGCATACCAAGGTGATTCGATGACCCACCGACCTGCTGCTGCTGCTGCTGCTG 1725
Db 668 AlaProGlySerLeuThrValAlaValValMetAspSerMetAspTrpPheAspProLys 687
QY 1726 -----GTGCGCAACAGTGGCGGAGTGTGCTGCGCAAGCAGGTGTGCGCGCGCGC 1776
Db 688 GluProGlyValAlaAlaAlaGlnIleSerLysLeuAsnArgAlaLeuLysProGlyGly 707
QY 1777 ATGTCATCTGCGCTCGCGCTCCCTCAGCCCGCTTACCGCGCTGATCCAGAGGCG 1836
Db 708 ArgValLeuLeuArgSerAlaGlyLeuLysProTrpTrpValGluGluPheAlaLysLeu 727
QY 1837 GGTCTTC-----GACGTGCGTGCATCCCGCGCGC 1866
Db 728 GlyPheAlaProLysArgValGlyCysArgSerProGluThrArgCysile----- 744
QY 1867 ACTCAGGCTTACATGGACCGCGTCAACATGTACAGCTCTTCTTACATGGCGCGCGGAG 1926
```

```
Db 745 -----AspArgValAsnMetTyrlleAlaSerCysTrpLeuLeuThrLysThr 759
QY 1927 GGC 1929
Db 760 Gly 760

RESULT 8
Q55SJ0_CRYNE PRELIMINARY; PRT; 820 AA.
ID Q55SJ0_CRYNE PRELIMINARY; PRT; 820 AA.
AC Q55SJ0_CRYNE PRELIMINARY; PRT; 820 AA.
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Hypothetical protein.
GN ORFNames=CNBE1930;
OS Cryptococcus neoformans var. neoformans B-3501A.
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Heterobasidiomycetes;
OC Tremellomycetidae; Tremellales; Tremellaceae; Filobasidiella.
OX NCBI_TaxID=283643;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=B-3501A.
RA Fung E., Hyman R.W., Rowley D., Bruno D., Miranda M., Fukushima M.,
RA Wickes B.L., Fu J., Davis R.W.;
RT "Cryptococcus neoformans serotype D sequencing.";
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AA01000024; EAL20831.1; -; Genomic_DNA.
KW Hypothetical protein.
SQ SEQUENCE 820 AA; 93328 MW; 0F694988B89353A2 CRC64;

Alignment Scores:
Pred. No.: 2,54e-37 Length: 820
Score: 881.00 Matches: 223
Percent Similarity: 43.6% Conservative: 110
Best Local Similarity: 29.2% Mismatches: 232
Query Match: 24.3% Indels: 198
DB: 2 Gaps: 21

US-10-620-914-44 (1-1947) x Q55SJ0_CRYNE (1-820)
QY 118 GGCAGCAAGAAGCGCGATGATCATCGCTGCGCTGAGAGCTTCTACGGCGCCAGGCC 177
Db 88 GlySerArgAspGlnLysAspHis-----LeuGluAlaPheTyrlleAlaGlyGlnAla 104
QY 178 GTCGCTTT----- 186
Db 105 AspLeuTyrlleAspThrThrArgSerGlnLeuLeuLysGlyArgGluThrMetLeuGlnLeu 124
QY 187 ---GTCGCGCGCTG-----GCCGAGCGCTG 210
Db 125 LeuAlaAlaHisLeuLysAlaGlnProMetValArgLeuProValSerAlaProSerLys 144
QY 211 AACCTCATCTGGTTCACCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 270
Db 145 ProArgileTrpValAspLeuGlyGlyThrGlyTrpAsnileGluLysMetAspGlu 164
QY 271 TACATCGACTGCGAAGTTCAAGTCCATCTACGTGCTGACCTGACCTGCTGCTGCTGCTG 330
Db 165 TyrLeuProLeuThrThrPheAspAlaileTyrlleileAspLeuLysGluProLeuLeu 184
QY 331 GAGTGGCGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 390
Db 185 GluValAlaArgAlaArgileLysAlaArgGlyTrpLysAsnValHisValLeuLysGln 204
QY 391 GAGCTTGGCAATTTTCGCGCGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 450
Db 205 AspAlaSerArgPheValLeuProGlu-----TrpGlu 215
QY 451 CTCACGATGATTCACCGCTTCCACACGTCATCGACCGAGCTTGTGCTGCTGCTGCTGCTG 507
```

```

Db 216 SerGlyAlaIleProPheTyrGlnValLeuAspArgCysAspGlnValLeuAspThr 235
QY 508 CAGACGCGCTGGTGGCGTGGAGTCTTACGTGAGC-----GGCAAGTACGAC 558
Db 236 GlnArgGlyLeuMetAlaValAlaAspPheTyrThrSerArgGluValGlyAsnLysGlu 255
QY 559 CTGCCCTG-----CGCAGATCCCTGGTTCGGCGCTTCTCTCGCGATCG 606
Db 256 ArgAlaIleGlyThrAlaSerLysArgValSerTrpPheSerLysTrpPheTrpGluCys 275
QY 607 ATCTTCACATCGACAACTTACATCGCGCGCGCGCTACCTGAGCAGCAAG 666
Db 276 TrpPheAspLeuAspGlyValHisLeuHisGlySerArgArgGluTyrLeuGluTyrLys 295
QY 667 CTGGAGCGCTGGGGAGCAGAACACCCAGGGTTTCGATCCCTACGTGCGCTGCGC 726
Db 296 MetGlyThrIleLysThrTyrAsnAlaArgAsnAsnPheLeuAsnThrTrpPheIleGln 315
QY 727 GCCCCTACTACGTGCTGATTGGC-----750
Db 316 IleProTyrTyrValPheLeuGlyCysSerArgGlnArgAspAlaSerAlaLys 335
QY 750 -----750
Db 336 SerPheThrLeuGluAlaGlyAsnArgLeuGlyGlnSerAspLeuGlyLeuLeuThrPro 355
QY 750 -----750
Db 356 ThrSerProPheThrAsnSerProSerValPheGlySerProSerProMetLeuGluMet 375
QY 750 -----750
Db 376 ProGluLeuValLeuGlyProSerAlaAlaGlnAlaThrGlnThrIlePheGluValGly 395
QY 751 -----CGCTTGGCCAGCGTTGGC 768
Db 396 AlaProLeuSerProPheHisTyrHisLeuArgLysAlaTrpArgIleProTyrLeuGlu 415
QY 769 CACGCGCTGCAGCAGAGCGGTGGAGCGCGCCCATGTTCCGCGCCACCTTCTCTGAC 828
Db 416 GluLysIleHisGluGlnPhe-----ArgThrHisIleTyr 427
QY 829 ACGCAGTCTGGGAGGACCGCGCGATATGAGGTGATGAGATCAACCCCAAGCAG 888
Db 428 GlyTrpThrTrpGluAspProAlaValAspValLysLysLeuAsnIleAsnLysAspAsp 447
QY 889 ACGGTGTGACCTGTAGCGCGGCTGCAATGCCCTGAAC---CTGCTGGTGCAGGG 945
Db 448 HisIleLeuAlaIleThrSerAlaGlyAspAsnValLeuHisTyrAlaLeuThrAlaLys 467
QY 946 GCGCGCAGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1005
Db 468 ProAlaArgIleHisAlaValAspMetAsnProCysGlnGlyHisIleLeuGluLys 487
QY 1006 AAGTGGCCATTACGACGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1065
Db 488 LeuAlaIleGlnAlaLeuGluTyrAsnAspPheTrpLeuIlePheGlyGluGlyArg 507
QY 1066 CACCGCGCATTTAGGAGCTTACGAGAAAGTGGCGCTTCTCTGTCGCAACACAGC 1125
Db 508 HisProGluPheAspLysLeuLeuThrThrLysLeuSerProPheLeuSerSerHisAla 527
QY 1126 CACAACCTTCTGGTCCAGCGCTCTGGTACTTCCAGCAGCGCGCTGTACTACACGCGGCG 1185
Db 528 TyrAlaTyrTrpLysThrHisThrSerGlnPheSerArgAsnPheTyrPheArgGlyTyr 547
QY 1186 ATGGGCAAGCTGTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1236
Db 548 SerGly-----TrpAlaLeuArgLeuAlaGlnIleAlaPhePheIleAlaGlyVal 564
QY 1237 GGCAAGACCGTCAAGCGCTCGCCAAACGCGCCCAATGGAGGAGCAGCGCGCTCTGTGG 1296

```

```

Db 565 ArgGlyAspValLysArgLeuCysGlnAlaThrSerThrAlaGluGlnGluArgIleTrp 584
QY 1297 GACAGCAACATGTCTATCCACTTTCGTGAAGACGGCGCGCTGGTGGTGGTGGTGGTGGTGG 1356
Db 585 GlnLysLysIle-----ArgPro---ValPheLeuAsn 594
QY 1357 GTCAGAGTTCGTGAGCTGGTGTCTTCAACAGCGCGTGTGGTTCGGCGCGCGCGGTG 1416
Db 595 LysValMetValLysLeuPheLeuGlyAsnProLeuPheAsnTrpHisAlaLeuGlyVal 614
QY 1417 CCGGCAAGCAGTACCGCTGATCAAGCGGAGCGCATCCCATTTGAGAACTACATCGCG 1476
Db 615 ProGlnAsnGlnMetAsnCysPheLeuGlnAspGly---SerValGluAspTyrValLys 633
QY 1477 CGCACCATGACCGCGTGGCGGAGACTCGCACGTGGCAGCAGCACTACTTCTACTAC 1536
Db 634 AlaThrLeuAspProIleProThrLeuSerThrLeuLysAspAsnTrpPhePheLeu 653
QY 1537 AACTGCTCACCGCGCAAGTTCCTGCGCAGCAACTGCGCCACCTACCTGCGCGAGCGGCC 1596
Db 654 LeuCysLeuAsnGlyArgTyrThrArgThrSerCysProAlaPheLeuLysProGluGly 673
QY 1597 TTCGCCACCTCAAG---AGTGGCGTGGTGAACAACCTGACCGCTCCCACTTCTTC 1653
Db 674 PheLysAlaLeuLysAsnSerLysSerThrAspAlaPheLysLeuHisThrAspThrIle 693
QY 1654 ATGAGGAGCTCAAGCG-----CGCACCTACACCAAGGTGATCTGTATGGAGCAC 1704
Db 694 LeuAsnValLeuArgGlyLeuProAspGluSerLeuThrLysIleValMetAspSer 713
QY 1705 GTGAGCTGGCTGAT-----ATGCCCGTGGCCAAAC---1734
Db 714 MetAspTrpPheAspProIleProGlyThrProLeuProGlnGlyAspSerThrAla 733
QY 1734 -----1734
Db 734 LeuAspThrLeuGlnAlaThrProGluLysAlaLeuGluHisLeuArgAlaGluLeuAsp 753
QY 1735 ---GAGTGGCGCGAGTCCCTGCGCCAGCAGGTGGCGCGCGCGCATCGTCTGGCGC 1791
Db 754 TyrGluIleLeuGlu---MetLysArgValLeuLysValGlyGlyIleAlaValTrpArg 772
QY 1792 TCGCGCTCTCCCTACGCGCGCTACCGCGCTATCCAGAAAGCGCGCTTCGACGTGCGC 1851
Db 773 SerAlaAlaLysArgProTyrArgGlnArgPheGluAlaGlyLeuLysValGln 792
QY 1852 TGCATCGC---CGCGCATCTAGCGCTACATGACCGCGCTCAACATGTACAGTCTCTTC 1908
Db 793 ProIleAspIleArgGluAsnGlnGluAlaIleAspArgValAsnMetTyrAlaSerPhe 812
QY 1909 TACATGCGC 1917
Db 813 TrpLysAla 815
RESULT 9
O6BK65 DEBHA
ID O6BK65 DEBHA PRELIMINARY; PRT; 763 AA.
AC O6BK65;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Similar to CA1783|IPF7635 Candida albicans IPF7635.
GN OrderedLocusNames=DEHAOP25982g;
OS Debaryomyces hansenii (Yeast) (Torulaspora hansenii).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Debaryomycetes.
OX NCBI_TaxID=4959;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=ATCC 36239 / CBS 767;
RX PubMed=15229592; DOI=10.1038/nature02579;
RA Dujon B., Sherman D., Fischer G., Durrans P., Casaregola S.,
RA Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talla E.,

```



```

Db 596 IleserGluAspAsnTyrPheTyrTyrLeuThrLeuLysGlyLysTyrThrProGlnAsn 615
QY 1570 TGCCCACTACTGCGCGAGCGGCTTCGCCACCCCTC-----1608
Db 616 CysProAspTyrLeuThrGluLysGlyTyrLysSerLeuThrThrLeuAsnArgLysSer 635
QY 1609 AAGAGTGGCTGTGGACACCTGACCGCTCCACC-----AAGTTCTTCATGGAG 1659
Db 636 LysGluAlaProIleAspAsnIleArgLeuHisThrAspMetLeuAsnAspValPheAla 655
QY 1660 GAGCTCAAGCGCGCCTACACCAAGGTATTCGTGATGACCACTGGAGTGGCTGGAT 1719
Db 656 ArgLeuSerLysLysSerLeuSerIleAlaIleMetAspHisMetAspTyrPheAsp 675
QY 1720 ATGCCGTGGCCAAAGAGTGGCGGAGTGGCTGCCAAGCAG-----GTTGGCGCGGC 1773
Db 676 ProAsnGlyGluAspAlaLeuAsnGluIleThrAlaLeuLysSerCysLeuAsnThrAsn 695
QY 1774 GGCATGCTACCTGGCGCTCCGCTCCCTCAGCCCGCCCTACCGGAGCTGATCCAGAAG 1833
Db 696 GlyArgValLeuLeuArgSerAlaSerThrAsnProTyrPyrIleLysThrPheGluAsp 715
QY 1834 GCGGCTTCACGCTGCGCTGCATCCGC---CGCGCCACTCAGGCTACATGACCGCGTC 1890
Db 716 LeuGlyPheThrCysLysAlaAlaGlyIleArgAspThrGlyIleSerIleAspArgIle 735
QY 1891 AACATGACAGCTCTCTTACATGGCCCGCGG 1923
Db 736 AsnMetTyrAlaSerThrTyrValCysThrLys 746

RESULT 10
Q7SGV5_NEUCR PRELIMINARY; PRT; 908 AA.
AC Q7SGV5;
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE Hypothetical protein (Related to S-adenosylmethionine:diacylglycerol
DE 3-amino-3-carboxypropyl transferase btaA).
GN Name=NCU03032.1; Synonyms=B19P24.050;
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=OR74A;
RA Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D.,
RA Jaffe D., FitzHugh W., Ma L.-J., Smirnov S., Purcell S., Rehman B.,
RA Elkins T., Engels R., Wang S., Nielsen C.B., Butler J., Endrizzi M.,
RA Qui D., Ianakiev P., Pedersen D., Nelson M., Washburne M.,
RA Selitrenikoff C.P., Kinsey J.A., Braun E.L., Zelter A., Schulte U.,
RA Kothe G.O., Jedd G., Mewes W., Staben C., Marcotte E., Greenberg D.,
RA Roy A., Foley K., Naylor J., Thomann N., Barrett R., Gnerre S.,
RA Kanal M., Kanysellis M., Mauceli E., Bielke C., Rudd S., Frishman D.,
RA Krystofova S., Rasmussen C., Metzenberg R.L., Perkins D.D., Kroken S.,
RA Cogoni C., Macino G., Catchside D., Li W., Pratt R.J., Osmani S.A.,
RA DeSouza C.C., Glass L., Orbach M.J., Berglund J., Voelker R.,
RA Yarden O., Plamann M., Seiler S., Dunlap J., Radford A., Aramayo R.,
RA Natvig D.O., Alex L.A., Mannheimaupt G., Ebbole D.J., Freitag M.,
RA Paulsen I., Sachs M.S., Lander E.S., Nusbaum C., Birren B.;
RT "The Genome Sequence of the Filamentous Fungus Neurospora crassa."
RT Nature 0:0-0(2003).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA Nyakatura G., Mewes H.W., Mannhaupt G.;
RL Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RA German Neurospora genome project;
RL Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.

```

```

CC CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AABX01000006; EAA36182.1; -; Genomic_DNA.
DR EMBL; BX842626; CAE7263.1; -; Genomic_DNA.
DR GO; GO:0008757; F:S-adenosylmethionine-dependent methyltransf. .; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000051; SAM bind.
KW Hypothetical protein; Methyltransferase; Transferase.
SQ SEQUENCE 908 AA; 101376 MW; C20AC06831AA4154 CRC64;

Alignment Scores:
Pred. No.: 9,76e-36 Length: 908
Score: 850.50 Matches: 222
Percent Similarity: 44.7% Conservative: 105
Best Local Similarity: 30.4% Mismatches: 235
Query Match: 23.5% Indels: 169
DB: 2 Gaps: 18

US-10-620-914-44 (1-1947) x Q7SGV5_NEUCR (1-908)
QY 205 CGCTCGAACCTCATCTGGGTGTGACCTGGGTGGCACTGGGAGAAATGTCGATATGATG 264
Db 178 ArgArgLysProIleTyrValAspValGlyGlyThrGlyTyrPheLysGluAlaMet 197
QY 265 GCTGATTACATCGACCTGCGCAAG---TTCAAGTCCATCTACGTGTGCGACCTGTGCCAC 321
Db 198 AlaLysPheValAsnValSerGluPhePheLysThrValTyrLeuValAspPheSerPro 217
QY 322 TCCTGTGGAGGTGGCCAAAGAAAGAGGCGAAGCCAGGGCTGGAGAGATGTCAGGTC 381
Db 218 SerLeuCysGluValAlaArgLysArgPheAlaArgLeuGlyTyrGluAsnValArgVal 237
QY 382 GTGGAGCGCGACGCTTGCCCAATTTGCG-----CCCTCGAGGGGCAAC 423
Db 238 IleCysThrAspAlaArgLysPheArgLeuGluAspTyrGluAspValAspGluGlyGlu 257
QY 423 -----423
Db 258 SerGlySerGlyAspSerSerProSerLeuSerGlyTyrTyrGlyGluThrLysProGly 277
QY 424 -----GGAGCGCTCATCCTTCTCTACTCGCTCAGCATGATCCACCGTTC 471
Db 278 ArgHisAlaGlyAlaGluLeuIleThrMetSerTyrSerLeuSerMetMetProAspTyr 297
QY 472 CACACGTCTACATCGACCGAGGCTTGTCTGTCTCTGCTCCAGCAGCGCTGGGGGTGGCC 531
Db 298 PheSerIleLeuAspSerLeuGluSerLeuLeuAlaProHisGlyLeuIleAlaValVal 317
QY 532 GACTTCTTACGTGAGCGGCAAGTACGAC-----CTGCCC 564
Db 318 AspPheTyrAlaGlnSerLysValAspPheThrPheArgAsnTyrThrGlyGlyLeuMet 337
QY 565 CTGGCGCAGATGCGCTGGTGGCGCGCTTCTTCTGGGATCGATCTTCGATCGACAC 624
Db 338 AsnArgHisValGlyTyrPheAlaArgAsnPheTyrArgSerTyrPheAspAlaAspArg 357
QY 625 ATTGACATCGCGCCGCGCGCTCTACTGAGCAGAGCTGGAGCGCGCTGG---681
Db 358 ValSerLeuGluProAlaArgAspTyrLeuGluTyrArgPheGlyThrValLeuThr 377
QY 682 -----GAGCAGAACACCCAGGGTTCGATCCCTACGTGCGCGTGG-----CTG 723
Db 378 ValAsnAlaArgAsnThrLeuGlyAlaIleProTyrTyrIleTyrLeuGlyCysLeu 397
QY 724 CGCGCCCTTACTACGTG-----741
Db 398 LysLysProPheSerThrSerSerLeuProHisGluIleValGluHisIleAspAlaIle 417
QY 741 -----741
Db 418 AlaThrGluSerProArgSerSerProArgLeuValGlyLysHisSerSerSerAlaThr 437

```

-!- CAUTION: The sequence shown here is derived from an EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is preliminary data.

EMBL; AABX01000006; EAA36182.1; -; Genomic_DNA.
 EMBL; BX842626; CAE7263.1; -; Genomic_DNA.
 GO; GO:0008757; F:S-adenosylmethionine-dependent methyltransf. .; IEA.
 GO; GO:0016740; F:transferase activity; IEA.
 InterPro; IPR000051; SAM bind.
 KW Hypothetical protein; Methyltransferase; Transferase.
 SQ SEQUENCE 908 AA; 101376 MW; C20AC06831AA4154 CRC64;

QY 742 -----TGGATTGGCCGC----- 753
 Db 438 AsnAlaLeuAlaPheAlaValGlyArgThrAlaProGluMetArgSerLysAlaPheAsn 457
 QY 754 -----CTGCCAGCGTTGGCCAGCGCCTGCAC 780
 Db 458 ThrAlaIleGluAsnIleSerAlaAsnLeuProLeuProSerPhePheTyrGlnAsnHis 477
 QY 781 -----GAGGAGCGGTGGAGCGGCCCATGTCTCCGCCACCTTC 822
 Db 478 HisTyrArgIleTyrTyrAspAspGlnLeuProLysHisThrGlnPheAsnAspGluTyr 497
 QY 823 CTGTACAGCGTGTGGAGAGCCCGAGCCGATGATGAGGTGATGAGATCAACCC 882
 Db 498 IleTyrAlaPheThrTyrGluAspSerArgValAspArgGluLeuLeuAsnLeuGlyPro 517
 QY 883 AAGCACAGTGTGACCTGTAGCGGCGGTGCAATCCCTGAACCTGCTGGTGCAG 942
 Db 518 AspAspValValLeuAlaIleThrSerAlaGlyAspAsnIleLeuSerTyrLeuMetGln 537
 QY 943 GGGCGCGCGAGGTGTGTGCTGCGACTGCAACCGCGCGAGTGGCGCTTCTGGAGCTG 1002
 Db 538 SerProAlaArgValHisAlaIleAspLeuAsnProAlaGlnAsnHisLeuLeuGluLeu 557
 QY 1003 AGAAGGTGGCCATTGACAGCTGGAGTTGAGAGCTGTGGCAGCTGTGGCGAGGC 1062
 Db 558 LysValAlaSerPheThrThrLeuAspTyrProAspValTyrLysIlePheGlyGluGly 577
 QY 1063 GTCACCCCGCATTTGAGGAGCTGACGAGAGAGCTGGCGCTTCTGTCGCAAC 1122
 Db 578 LysHisProAspPheArgSerLeuLeuIleSerLysLeuSerProHisLeuSerGlyArg 597
 QY 1123 AGCCACAACTTCTGGTCCAGCGCTCTGTACTTCCAG-----CACGGCGCTG 1170
 Db 598 AlaPheGlnTyrTrpLeuSerAsnAlaHisIlePheThrAspProAlaGlyArgGlyLeu 617
 QY 1171 TACTACAGCGCGGCTGAGCAAGCTGTGTGGTGTGCTGAGTGTGGCGCTGTGGTGTG 1230
 Db 618 TyrAspThrGlyGlySerArgTyrAlaIleArgPhePheArgTyrPheThrLeuPhe 637
 QY 1231 GGACTGGCAAGACCTCAAGCGCTCGCCAGCGCCACATGAGGAGGAGCGCGCT 1290
 Db 638 PheCysArgSerAlaValArgArgLeuLeuSerThrProThrLeuGluGlyGlnArgSer 657
 QY 1291 CTGTGGAGACAACTCATCTCATCTCGTGAAGAAAGCGCGCGCTGTGTGTG 1350
 Db 658 IleTyrHisThrLysIle-----ArgProCys----- 666
 QY 1351 CTGTTCTGCTCAAGTGTGTGAGC---CTGTGCTCTTCAACAGCGCGTGTGTGTGCGC 1407
 Db 667 LeuLeuAsnArgPheValAsnGlyLeuValLeuSerSerAspAlaPheLeuTyrSerAla 686
 QY 1408 GGGCGGTGCGCGGAGCAGTACGCTGATCAAGCGGAC----- 1449
 Db 687 LeuGlyValProLysAsnGlnValAlaMetIleGluAlaAspTyrHisArgArgSerIle 706
 QY 1450 -----GGCATCCCATTTGAGAAC 1467
 Db 707 SerSerSerThrThrProSerSerLysGluLysProSerArgAlaGluAlaIleLeuHis 726
 QY 1468 TACATCGCGCGACCATGAGCGGTGGGAGAACTCGCAGCTGGCGCAAGCAAACTAC 1527
 Db 727 TyrThrThrSerThrLeuAspProValLeuSerThrSerHisLeuAlaSerAspAsnPro 746
 QY 1528 TTCTACTACAACTGCTCCCGGCAAGTCTCTCGCGGACAACTGCGCCACCTACCTCGC 1587
 Db 747 TyrTyrLeuValCysValLeuGlyGlnTyrThrArgGlnCysHisProAspTyrLeuSer 766
 QY 1588 GAGCGCGCTTCCGCCCTCCAGAGT---GGCGTGTGGCAACCTGACCGCTCCACC 1644
 Db 767 ProAlaAlaHisSerIleLeuSerAlaProGlyAlaPheAspGlyLeuArgIleHisThr 786
 QY 1645 AACTTCTTATGAGGAGCTC-----AAAGCGCGCACCTTACACCAAGGTGATTCTG 1695

Db 787 AspGluIleGlnGluValLeuAlaArgPheGlnProGlyThrLeuThrValAlaVal 806
 QY 1696 ATGGACACAGTGGCTGGATATGCC----- 1725
 Db 807 MetAspSerMetAspTrpPheAspProProSerProGluGluGluGlyArgGly 826
 QY 1726 GTGGCCACAGAGTGGCCGAGTGCCTGGCCAGCAGTTCGGCGCGCGCATCGTCATC 1785
 Db 827 LysAlaArgGluGlnValArgArgLeuAsnArgAlaLeuLysValGlyGlyValLeu 846
 QY 1786 TGGCGCTCGCTCCCTCCTAGCCCGCTAGCCGAGCTGATCCAGAGCGCGCTTCGAC 1845
 Db 847 LeuArgSerAlaGlyValGluProTyrTyrValArgValPheValGluGluGlyPheGly 866
 QY 1846 GTGCGCTGATC-----CGCGCGCCACTCAGGCTACATGAGCCGCGTC 1890
 Db 867 AlaArgArgValGlyCysArgGluSerGlyArgGlyAspGlnGluCysIleAspArgVal 886
 QY 1891 AACATGTACAGCTCTTCTACATGCGCCCGCG 1923
 Db 887 AsnMetTyrAlaSerCysTrpIleLeuGluLys 897
 RESULT 11
 Q5KGX9_CRYNE PRELIMINARY; PRT; 797 AA.
 ID O5KGX9; 30, Created)
 AC O5KGX9;
 DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
 DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
 DE Hypothetical protein.
 GN ORFNames=CNE01970;
 OS Cryptococcus neoformans var. neoformans JEC21.
 OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Heterobasidiomycetes;
 OC Tremellomycetidae; Tremellales; Tremellaceae; Filobasidiella.
 OX NCBI_TaxID=214684;
 RN [1]
 RC NUCLEOTIDE SEQUENCE.
 RP STRAIN=JEC21;
 RA Loftus B., Amedeo P., Roncaglia P., Vamathevan J., Utterback T.,
 Van Aken S., Fraser C.; to the EMBL/GenBank/DBJ databases.
 RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
 RC STRAIN=JEC21;
 RX PubMed=15653466; DOI=10.1126/science.1103773;
 RA Loftus B.J., Fung E., Roncaglia P., Rowley D., Amedeo P., Bruno D.,
 Vamathevan J., Miranda M., Anderson I.J., Fraser J.A., Allen J.E.,
 Bosdet I.E., Brent M.R., Chiu R., Doering T.L., Donlin M.J.,
 D'Souza C.A., Fox D.S., Grinberg V., Fu J., Fukushima M., Haas B.J.,
 Huang J.C., Janbon G., Jones S.J.M., Koo H.L., Krzywinski M.I.,
 Kwon-Chung K.J., Lengeler K.B., Maiti R., Marra M.A., Marra R.E.,
 Mathewson C.A., Mitchell T.G., Pertea M., Riggs F.R., Salzberg S.L.,
 Schein J.E., Shvartsbeyn A., Shin H., Shumway M., Specht C.A.,
 Suh B.H., Tenney A., Utterback T.R., Wickes B.L., Wortman J.R.,
 Wye N.B., Kronstad J.W., Lodge J.K., Heitman J., Davis R.W.,
 Fraser C.M., Hyman R.W.;
 RT "The genome of the basidiomycetous yeast and human pathogen
 RT Cryptococcus neoformans";
 RL Science 307:1321-1324 (2005).
 DR EMBL; AB017345; AAM43531.1; -; Genomic DNA.
 KW Complete proteome; Hypothetical protein.
 SQ SEQUENCE 797 AA; 90388 MW; FD1082B7EB2B240 CRC64;

Alignment Scores:
 Pred. No.: 1.73e-35 Length: 797
 Score: 845.50 Matches: 221
 Percent Similarity: 42.8% Conservative: 104
 Best Local Similarity: 29.1% Mismatches: 220
 Query Match: 23.4% Indels: 215
 DB: 2 Gaps: 21

US-10-620-914-44 (1-1947) x Q5KGX9_CRYNE (1-797)


```
QY 1015 ATTACGAGCTGGAGTTGAGGAGCTGTGCGACGCTGTTGCGGAGGCGGTGCACCGCGC 1074
Db 477 PheMetAlaLeuGlyHisArgAspValTrpLysPheGlyGluGlyHisProGlu 496
QY 1075 ATTGAGGAGCTGACGAGAAAGCTGGCGCCCTTCCTGTCGCAAAACAGCACCAACTTC 1134
Db 497 PheArgGluLeuLeuLeuSerArgLeuSerAlaHisLeuSerGlnAlaPheGlnTyr 516
QY 1135 TGTCCAAAGCGCTCTGGTACTTCCAG-----CACGGCTGTACTACAGGGC 1182
Db 517 TrpLeuGluHisThrHisIlePheThrSerLysTyrGlyLysGlyLeuTyrGluThrGly 536
QY 1183 GGCATGGCAAGCTGTGCTGGGTCTGACGTGCTGCGCGTGTGCTGGACTGGGCAAG 1242
Db 537 GlySerArgHisAlaLeuLysMetValArgTyrLeuPheLysValPheGlyLeuGluGly 556
QY 1243 ACCGTCAAGCGCTCGCCCAAGCGGCCCAATGAGGAGGAGGAGCGCGCTGTGGGACAGC 1302
Db 557 GlnValLysLysLeuCysGluAlaGlnThrLeuAlaGluGlnArgGluIleTrpProLys 576
QY 1303 AACATGCTCATCACTTCGTGAAGACGGCGCCCAAGCGCTGTGTGGCTGTTCGTCAAG 1362
Db 577 IleArgAlaValLeuMet-----SerLysProLeuHisTrp----- 588
QY 1363 TTCGTGAGCTGTGCTCTTCAACAGGCGCTGTGTGGTTCGGCGGCGGTGCGCGGC 1422
Db 589 -----AlaValValSerThrGluTrpPheAlaTrpLysAlaAlaGlyValProArg 605
QY 1423 AAGCAGTACGCGCTGATCAAGGGGAC----- 1449
Db 606 AsnGlnArgAsnMetIleValAspAspTyrPheLysArgLeuGlyLeuThrLysAspMet 625
QY 1450 -----GGCATCCCATGAGAACTACATCGCGCGCCACCATGGAC 1488
Db 626 AsnGlnGlyLysAspIleSerGlyArgSerIleTrpGlnTyrValValAspThrLeuAsp 645
QY 1489 GCGTGGCGGAGAACTCGCAGCTGCGGACGAGAACTTCTTACTACAACTGCTCACC 1548
Db 646 ProValValAsnGluThrMetIleSerAsnAspAsnTyrPheTyrPheLeuCysLeuGln 665
QY 1549 GCGAAGTCTTCGCGGCAAACTGCGCCACCTACCTGCGGAGGCGCGCTTCGCGCACCTC 1608
Db 666 GlyLysPheSerArgArgCysHisProThrTyrLeuSerProGlnAlaHisValLysLeu 685
QY 1609 AAGAGT---GCGGTGGTGAACACTGACCTGCTCCACCACTTCTTCATGGAGGACTC 1665
Db 686 SerSerProGlyAlaPheAspGlyLeuArgIleHisThrAspGluIleAsnGluValIle 705
QY 1666 AAA-----GCGCGCACTACCAAGGTGATTCTGTGGACCACTGCGGCTGGCTG 1716
Db 706 LysArgIleThrProArgSerLeuThrIleAlaValIleMetAspSerMetAspTrpPhe 725
QY 1717 GATATGCCGCTGGCCAAACGAGCTG-----GCCGAGTGCCTGGCCAAAGAGTTCG 1767
Db 726 Asp---ProGluGlyThrGluAlaSerIleGlnAlaGlnLysLeuAsnHisAlaLeuLys 744
QY 1768 CCGGCGCGCATGTCATCTGCGCTGCTCCCTGACCGCGCCCTACGCGGAGCTGATC 1827
Db 745 MetAspGlyArgIleLeuLeuArgSerAlaSerIleGluProTrpTyrIleLysGlnPhe 764
QY 1828 CAGAAGCGGCTTCGAGCTGCGCTGCATC---CGCGCGGCACTCAGGGGTACATGAC 1884
Db 765 GluGlnAsnGlyPheThrAlaArgValGlyAlaArgPheProGlySerCysIleAsp 784
QY 1885 CGCGTCAACATGTACAGCTCTTCTACATGCGCGCGCGG 1923
Db 785 ArgValAsnMetTyrAlaSerThrTrpIleCysThrLys 797
RESULT 13
Q5B0L6 EMENI
ID Q5B0L6 EMENI PRELIMINARY; PRT; 790 AA.
AC Q5B0L6,
```

```
DT 10-MAY-2005 (TrEMBLrel. 30, Created)
DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE Hypothetical protein.
GN ORFNames=ANS914.2;
OS Aspergillus nidulans FGSC A4.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eukotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=227321;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=FGSC A4;
RA Birren B., Nusbaum C., Abouelleil A., Allen N., Anderson S.,
RA Arachchi H.M., Barna N., Bastien V., Bloom T., Boguslavskiy L.,
RA Boukhalter B., Butler J., Calvo S.E., Camarata J., Chang J.,
RA Choepel Y., Collymore A., Cook A., Cooke P., Corum B., Dearellano K.,
RA Diaz J.S., Dodge S., Dooley K., Dorris L., Elkins T., Engels R.,
RA Erickson J., Faro S., Ferreira P., Fitzgerald M., Gage D., Galagan J.,
RA Gardyna S., Gnerre S., Graham L., Grand-Pierre N., Hafez N.,
RA Hagopian D., Hagos B., Hall J., Horton L., Hulme W., Iliev I.,
RA Jaffe D., Johnson R., Jones C., Kamat M., Kamat A., Karatas A.,
RA Kells C., Landers T., Levine R., Lindblad-Toh K., Liu G., Lui A.,
RA Ma L.-J., Mabbitt R., Maclean C., Macdonald P., Major J., Manning J.,
RA Matthews C., Mauceli B., McCarthy M., Meldrim J., Meneus L.,
RA Mihova T., Mlenga V., Murphy T., Naylor J., Nguyen C., Nicol R.,
RA Nielsen C.B., Norbu C., O'Connor T., O'Donnell P., O'Neill D.,
RA Oliver J., Peterson K., Phunkhang P., Pierre N., Purcell S.,
RA Rachupka A., Ramasamy U., Raymond C., Retta R., Rise C., Rogov P.,
RA Roman J., Schauer S., Schuback R., Seaman S., Severy P., Smirnov S.,
RA Smith C., Spencer B., Stange-Thomann N., Stojanovic N., Stubbs M.,
RA Talanas J., Tesfaye S., Theodore J., Topham K., Travers M.,
RA Vassiliev H., Venkataraman V.S., Viel R., Vo A., Wang S., Wilson B.,
RA Wu X., Wyman D., Young G., Zainoun J., Zembek L., Zimmer A., Zody M.,
RA Lander E.;
RT "Genome Sequence of Aspergillus nidulans.";
RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AACD0100101; EAA5777.1; -; Genomic DNA.
DR GO; GO:0004719; F:protein-L-isoaspartate (D-aspartate) O-meth. .; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006464; P:protein modification; IEA.
DR InterPro; IPR00682; PCMT.
DR InterPro; IPR000051; SAM_bind.
KW Hypothetical protein.
SQ SEQUENCE 790 AA; 90079 MW; C3D808C2EC30CC7B CRC64;

Alignment Scores:
Pred. No.: 1.18e-30 Length: 790
Score: 752.00 Matches: 211
Percent Similarity: 42.3% Conservative: 107
Best Local Similarity: 28.1% Mismatches: 251
Query Match: 20.8% Indels: 182
DB: 2 Gaps: 22

US-10-620-914-44 (1-1947) x Q5B0L6 EMENI (1-790)
QY 100 CTGCGCCATATGTGTTGCGGACGAGAGGCGGATGATCAGCTGCTCGCTGGAGAGC 159
Db 58 LeuArgPheIleTyrAlaSerPheLeuLysPheHisAspLysGlyCysGlnLeuThrAla 77
QY 160 TTCTACGCGCCGACGCGCTGCTT----- 186
Db 78 -----ProGlnAlaThrValTyrAspAlaThrArgLysArgLeuLeuArgGlyArg 94
QY 187 -----GTCGCCGCGCTG-----GCC 201
Db 95 GluAspMetLeuGlyLeuValAlaAlaGlnLeuLysTyrLysValGluAsnLysGluLeu 114
QY 202 GAGCGCTCGAAACCTCATCTGGGTTCACCTGGGTGGTGGCTGGGAG-----AATGTC 255
Db 115 ArgLysGlyLysAlaIleTrpValAspIleGlyGlyThrGlySerArgTyrAsnIle 134
QY 256 GATATGATGCTGATTCATCGACCTGGCGAAGTTCAAGTCC---ATCTACGTGTCGAC 312
Db 256 GATATGATGCTGATTCATCGACCTGGCGAAGTTCAAGTCC---ATCTACGTGTCGAC 312
```

135 GluAlaMetAlaGluPheValProGluPhePheSerHisValTyrLeuValAsp 154
QY 313 CTGTGCCACTCGCTGTGCGAGGTGGCCAAAGAGCGCAAGGCGGCTGGAAGAT 372
Db 155 LeuSerProSerLeuLeuGluValAlaArgGlnArgPheGluArgLeuGlyTrpLysAsn 174
QY 373 GTCCAGTGTGTGGAGCGGAGCTGTGCAATTTGCGCCCTGTAG 417
Db 175 ValThrValValCysGlnAspAlaArgAlaPheGlnLeuProGluAspHisIleAspPro 194
QY 418 -----GGCACCGGACGCTCATCACTTCTCTACTCGCTCACGATGATT 462
Db 195 LeuLysSerValGlyAlaGlyAlaAspLeuValThrMetSerTyrSerLeuSerMetIle 214
QY 463 CCACCGTTCCACACGATCGACGAGCTGTGCTGTACCTGTCCCAAGACGCGCTGGTG 522
Db 215 ---ProValGlnSerIleValAspValSer----- 223
QY 523 GCGTTCGCCACTTCTACGTGAGCGGCAAGTACGACCTGCCCTGCGCCAGATGCCCTGG 582
Db 224 ---SerArgAsnTyrIleGlyGlyValPheAsn-----ArgHisValAsnTrp 238
QY 583 TCGCGCGTTCCTCTCGGATCGATCTTCGACATCGCAACATTTGACATCGCGCCCGAG 642
Db 239 LeuGlyArgAlaPheTrpArgAlaTrpPheGluAlaAspArgValAsnLeuAspAlaAla 258
QY 643 CGCGCGCCCTACTGGAGCAGAGCTGGAGCGGTGTGGAGCAGACACCCAG----- 696
Db 259 ArgArgAspTyrLeuGluTyrArgPheGlyThrValIleSerAlaSerGluArgAsnTyr 278
QY 697 -----GGTTCGATCCCTAGTGCCTGCTGCGCGCCCTCTACTAGTGTGATGGC 750
Db 279 LeuLeuGlyGlyIle-----ProTyrTrpIlePheIleGly 753
QY 751 CGC----- 753
Db 291 ArgGlnLysAspIleTyrProAsnGlnAlaSerArgGluAlaIleGluLysLeuAspAla 310
QY 753 ----- 753
Db 311 SerPheThrGluSerProTyrLeuSerProAlaAsnHisArgGluGluMetAsnAla 330
QY 753 ----- 753
Db 331 IleIleGlnSerThrGlnGluIleHisSerLysAlaTyrGluSerAlaValValAsnLeu 350
QY 754 -----CTGCCCGCGGTGGCCAGCGCTGCGACGCGGAGCGCGT----- 792
Db 351 SerAlaAsnLeuProLeuProAlaAlaPheTyrGlnAsnHisIleTyrArgIlePheTyr 370
QY 793 -----GAGCGCGCGCCCATGTTCGCCGCCACCTTCTCTGACAGCAGTGTGG 840
Db 371 AsnAspLeuLeuProLysHisThrGlnPheLysAsnGluTyrIleTyrAlaPheAsnTrp 390
QY 841 GAGACCCCGAGCGGATATGAGGTGATGAGATCAACCCCAAGACGACGCGTGTGACC 900
Db 391 GluAspProArgValAspHisArgLeuLeuAsnIleLysArgAspValIleLeuAla 410
QY 901 CTGACTAGCGGCGCTGCAATGCCCTGCACTGCTGTGCGAGGGCGCGCGAGGTG 960
Db 411 IleThrSerAlaGlyAspAsnIleLeuAspTyrLeuGlnLysSerProArgValHis 430
QY 961 TCGGTGAGTCAACCCCGCGAGTGTGCGCTTCTGAGCTGAAGAGGTGGCCATTTCAG 1020
Db 431 AlaValAspLeuAsnProAsnGlnAsnHisLeuIleGluLeuLysValAlaSerPheIle 450
QY 1021 CAGCTGAGTTTGAAGACGTGTGGAGCTTCTGCGAGGCGGTGACCGCGCATTTGAG 1080
Db 451 AlaLeuGlyHisArgAspValTrpLysIlePheGlyGlyLysHisProGluPheArg 470
QY 1081 GAGCTGTACGAGAGAGTGGCGCTTCTGTCGCAACACGACCACTTCTGCTCC 1140
Db 471 AsnLeuLeuIleSerArgLeuSerProHisLeuSerSerGlnAlaPheGlnTyrTrpLeu 490

1141 AAGCGCTCTGTGTACTTCCAG-----CACGGCTGTACTTACCAGGCGCGCATG 1188
QY 491 GluHisSerHisValPheThrSerSerSerGlyArgGlyLeuTyrGluThrGlyGlySer 510
Db 1189 GGCAGCTGTGTGGTGTGTGTCAGTGCCTGCGCGTGTGTGCTGGACTGGCAAGACGCTC 1248
QY 511 ArgHisAlaIleLysMetIleArgTyrIlePheLysValPheGlyLeuGluGlnVal 530
Db 1249 AAGCGCTCTGCCAACCGCCCAACATGAGAGCAGCGCGCTCTGTGGGACAGCAACATG 1308
QY 531 LysLysLeuCysGluAlaGlnThrLeuAlaGluGlnArgGlnIleTrpProArgIleArg 550
Db 1309 CTCACTCACTTCTGTGAAGAACGCGGCCAACCGCTGTGTGCTGTTCTCAAGTTCGTG 1368
QY 551 AlaValLeuLeu-----SerLysProLeuHisTrp----- 560
QY 1369 AGCTGTGTCTCTTCAACAGCGGTGTGTGTGCTGCGCGCGCGTGGCGGCGACAG 1428
Db 561 ---AlaValValSerThrGluTrpPheAlaTrpLysAlaAlaGlyValProArgAsnGln 579
QY 1429 TAGCGCTGATCAAGCGGAC----- 1449
Db 580 ArgAsnMetIleValGluAspTyrPheLysArgAsnGlyLeuThrLysAspMetLysAsn 599
QY 1450 -----GGCATCCCATTTGAACTACATCGCGCGCACCATGAGCGCGTGT 1494
Db 600 AlaLysAspIleSerGlyLysSerIleTyrGluTyrValValAspThrLeuAspProVal 619
QY 1495 GCGGAGAACTCGACGCTGCGCAGCAGAACTACTTCTACTACAACTGCTCACCAGCAAG 1554
Db 620 ValHisGluThrMetIleSerAsnAspAsnTyrPheTyrTyrLeuLysLeuLysGlyGln 639
QY 1555 TTCTCTGCGGACAACTGCCCACTACCTGCGGAGCGCGCTTCCGACCCCTCAAGAGT 1614
Db 640 PheSerLysArgCysAlaProAlaTyrLeuSerProGlnAlaHisValLysLeuSerSer 659
QY 1615 ---GGCGTGTGTGAACTGACCTGCTCCACC-----AACTTCTTCATGAGGAG 1662
Db 660 ProGlyAlaPheAspGlyLeuArgIleHisThrAspGluIleAsnGluValIleGlyArg 679
QY 1663 CTCAAAGCGGACCTACACCAAGGTGATCTGATGACCACTGAGCTGCTGGATATG 1722
Db 680 IleThrProArgSerLeuThrIleAlaValIleMetAspSerMetAspTrpPheAsp--- 698
QY 1723 CCGTGTGCGCAACGAGCTG-----GCGAGTGTGCGCAAGCAGGTTCGCGCGGC 1773
Db 699 ProGluGlyThrGluAlaThrAlaGlnAlaGlnLysPheAsnGlnAlaLeuLysMetAsp 718
QY 1774 GGCATGCTCATCTGGCGCTCCGCTCCCTCAGCCCGCCCTACCGCGGAGTGTATCCAGAAG 1833
Db 719 GlyArgIleLeuLeuArgSerAlaSerIleGluProTyrTrpTyrIleLysAsnPheGluGlu 738
QY 1834 GCGGCTTTCAGCTGCGCTGCATCCCGCGCGCCACTCAGGCG---TACATGGACCGCGTC 1890
Db 739 AsnGlyPheSerAlaArgValGlyAlaArgPheSerGlyThrCysIleAspArgVal 758
QY 1891 AACATGTACAGTCTCTTACATGGCGCGCG 1923
Db 759 AsnMetTyrAlaSerThrTrpIleCysThrLys 769
RESULT 14
Q7UYV8 RHOB
ID Q7UYV8 RHOB PRELIMINARY; PRT; 415 AA.
AC Q7UYV8
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN OrderedLocusNames=RB360;
OS Rhodospirillum rubrum
OC Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;
OC Planctomycetaceae; Pirellula.

```

OX NCBI_TaxID=117;
RN [1];
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=1;
RA MEDLINE=22735913; PubMed=12835416; DOI=10.1073/pnas.1431443100;
RX Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,
RA Ludwig W., Gade D., Beck A., Borzym K., Heitmann K., Rabus R.,
RA Schleener H., Amann R., Reinhardt R.;
RT "Complete genome sequence of the marine planctomycete Pirellula sp.
RL strain 1.";
RT Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303 (2003).
DR EMBL; BX294133; CAD71533.1; -; Genomic DNA.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 415 AA; 47529 MW; 88D3194F7446D252 CRC64;

Alignment Scores:
Pred. No.: 9,09e-16 Length: 415
Score: 463.00 Matches: 124
Percent Similarity: 48.1% Conservative: 79
Best Local Similarity: 29.4% Mismatches: 158
Query Match: 12.8% Indels: 62
DB: 2 Gaps: 13

US-10-620-914-44 (1-1947) x Q7UYV8_RHOBA (1-415)
QY 742 TGGATTGGCGCGCTCCAGCGTTGGCCACGCGTGCACGAGGAGCGCGTGGAGCGCG 801
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 5 TrpLeuGlyAsn-----LysCysPheLysValValHisGlnLysAsnLeu----- 19
QY 802 CCCATGTTCCCGCCACCTTCTGTACACGAGTCGTGGGAGGACCCCGAGCGGATATG 861
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 20 -----ValTyrAsnThrCysTrpGluAspProArgLysAspArg 32
QY 862 GAGTGATGAGATCAACCCAGGACACGGTGTGACCTGACTAGCGCGGCTGCAAT 921
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 33 GlnAlaLeuSerLeuGlyProAspAspSerValLeuValIleThrSerAlaGlyCysAsn 52
QY 922 GCCTGAACTGCTGGTGGAGGGCGCGCAGTGTGCTGGTGGAGTGCACCCCGCG 981
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 53 AlaLeuAspTyrAlaLeuGlnAlaProLysSerValHisAlaValAspMetAsnProLeu 72
QY 982 CAGTCGGCGCTTCTGGAGCTGAAGAAGTGGCCATTCAGCAGCTGGAGTTTGAGGACGTG 1041
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 73 GlnAsnAlaLeuLeuGluLeuLysCysAlaSerIleArgCysLeuThrTyrAspAspPhe 92
QY 1042 TGGCAGCTGTTGGCGAGGGGTGCACCCGCGCATTCAGGAGCTGTACGAGAGAACGCTG 1101
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 93 PheAsnValPheGlyArgGlyTyrHisProAlaTrpGlnSerLeuTyrHisLysHisVal 112
QY 1102 GCGCCCTTCTGTCGCAAAACCGACCCACACTTCTGGTCCAGCGCTCTGTACTTTC--- 1158
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 113 ArgGlyAlaLeuThrProAspThrArgAlaIleTrpAspArgArgLeuAspPheAsp 132
QY 1159 -----CAGCAGCGCTGTACTACAGGGCGGCGATGGCAAGCTGTGCTGGGTG 1206
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 133 GlyThrSerArgArgLysSerPheTyrPheArgGlyThrSerGlyLeuPheAlaTrpLeu 152
QY 1207 CTGCAG---TGCTTCGGCGGTGCTGCTGGAGTGGCGAAGACCGTCAAGCGCTCGCCAC 1263
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 153 ValAsnGlyTyrLeuLysArgProAlaGlyLeuArgGluAlaIleAlaGluLeuGln 172
QY 1264 GCGCCCAACATGAGGAGGACGCGCTGTGGGACAGCAACATGCTCCACTTCGTGCT 1323
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 173 AlaGluSerValAspGlnArgGluIleTyrGlnLysArgGluIleAsnArgLeuLeu 192
QY 1324 AAGAACGGCCCAAGCCGCTGGTGTGCTGCTCAAGTTTCGAGCCTGGTCTTTC 1383
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 193 TrpSer-----LysProLeuArgTyrAlaLeuArgArgAspThrThrLeuAlaMetLeu 210
QY 1384 -----AACAGCCGCTGCTGTGTTGCTGGCGGGCGGCTCCCGGC 1422
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 211 GlyValProArgSerGlnArgAsnGlnLeuAspGlnCysTyrProGlyGlyGlyGly 230

```

RESULT 15

Q4NR00_9DELTA PRELIMINARY; PRT; 767 AA.

AC Q4NR00; 13-SEP-2005 (T-EMBLrel. 31, Created)
 DT 13-SEP-2005 (T-EMBLrel. 31, Last sequence update)
 DT 13-SEP-2005 (T-EMBLrel. 31, Last annotation update)
 DE Liga.
 GN ORFNames=AdehdRAFT_0943;
 OS Anaeromyxobacter dehalogenans 2CP-C.
 OC Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
 OC Cystobacterineae; Myxococcaceae; Anaeromyxobacter.
 OX NCBI_TaxID=290397;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=2CP-C;
 RG US DOE Joint Genome Institute (JGI-PGF);
 RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,
 RA Hammon N., Israni S., Pittluck S., Richardson P.;
 RT "Sequencing of the draft genome assembly of Anaeromyxobacter
 dehalogenans 2CP-C.";
 RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=2CP-C;
 RG US DOE Joint Genome Institute (JGI-ORNL);
 RA Larimer F., Land M.;
 RT "Annotation of the draft genome assembly of Anaeromyxobacter
 dehalogenans 2CP-C.";
 RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
 CC -!- CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 CC preliminary data.

Qy	1138	ACCAGAAGTTGTGGC-----TGGTTTGCACAGGAAGGGCGGCAGCTTCTTCTCGTACA	1088
Db	303	-----AlaCysGlyArgArgTTPGlyAlaAlaTrp-----ProAlaThrAlaAarg	318
Qy	1084	GCTCCTCAATGCGGGGTGCACGCCCTCGCCGAACAGCTGCCACACGTCCTCAAACCTCCA	1025
Db	319	AlaProThrArgCysGlyAlaAargSerProAargThrIleAlaSerArgArgCysAlaGly	338
Qy	1024	GCTGCTGAATGGCCACCTTCTTCAGCTCCAGAAAGCCGACTCGCGGGGTTCGAGT---	968
Db	339	ArgProGlyTrp-----ArgAlaAlaAargArgAlaAargCysAlaSerGlyCysSerArg	356
Qy	967	-----CCACGGACACACCTGGC---	950
Db	357	SerArgArgSerAlaCysSerAlaAargTrpArgAlaThrProAargArgGlyProGlyPro	376
Qy	949	-----CGGCCCTCTGCACCA-	935
Db	377	CysAlaSerAlaAargSerProSerProAargArgArgSerThrArgTrpThrGlyTrpThr	396
Qy	934	---GCAGGTTTCAGGCATTCAGCCCGCGCTAGTCAGGGTCAGCACCCGTCCTCT---	884
Db	397	ThrArgGlyProGlyArgCysGlyProAargAlaSerProAlaGlyProProProSer	416
Qy	883	-----TGGGGTTGATCTCATCACCT-----CCATATCCGGCTCGG	848
Db	417	ArgArgSerGlyThrSerProAargProAargProAlaGlyArgSerSerArgProAargSer	436
Qy	847	GGTCTCCACGACTCGGT-----ACAGGAAGTGGCGG	812
Db	437	GlyProProAargAlaCysProCysGlyThrArgThrArgCysSerArgTrpPro	456
Qy	811	GGAACTGGCGCGCGCTCCA-----CGCGCTCTCTCGT-----	779
Db	457	ArgProAargAlaCysProProAargTrpArgSerHisAlaProAargAspValProPro	476
Qy	779	-----	779
Db	477	ArgAspArgProAlaHisAlaAlaAargHisGlyAlaProAlaHisArgGlyGlyAla	496
Qy	778	---GCAGGGCGTCGCCAAGC-----TGGCGAGCGCGCCCAATCCACACGT	737
Db	497	GlyGlyGlyArgGlyAlaAargValProProLeuAlaAlaHisGlyProAlaAlaAla	516
Qy	736	AGTAGGGGCGCGCAGCCACGCA-----CGTAGGGGATCGAAC-----	698
Db	517	GluaGlyAspGlyAlaGlyAlaAspProAlaValAargGlyGlyAspHisHisHis	536
Qy	697	-----CCTGGGTCTTCGTCCC	680
Db	537	GlyAspArgArgGlnAargGlyGluGlyValAargProGlyGlyGlnLeuAargPro	556
Qy	679	ACAGGCGCTCCAGCTTCTGCTCCAGTAGCGC-----	647
Db	557	ValProLeuAargAlaGluGlyProAlaAargArgGlyGlyAlaValProHisHisAargPro	576
Qy	646	-----GGCGCTCGGGCCGATCAATGTGTTCGATGTGGAAGATCGATCGCCAGAAGA	593
Db	577	GlyHisGlyValAargSerAaspAlaGlyAlaAargAlaHisAargAlaAargGlyVal	596
Qy	592	AAACGGCGCACCAAGGCCA-----TCTGGCGCAGGGCGAGTCTGCTACT	551
Db	597	GlyGlyAlaGlyAargAlaAargProAlaProGlyAargGlyAlaGlyAargGlyAargAla	616
Qy	550	TGGCGCTCAGTAAAGTGGCAACGCCCAAGGCGGTCTGGGACAGTACGACGACAG	491
Db	617	ArgGlyAargAargAlaAargAargAargAlaAarg---GlyGlyGlyAlaAlaAla	635
Qy	490	CCTGGTCGATGACGTTGTGGAACGGTGGAAATCATCTGTAGCGAGTAGGACGAAGTGATGA	431
Db	636	AlaAargArg-----ArgValAlaGluValAlaValAargAargSerArgAlaAla	651
Qy	430	GCCTCGCGGTGCCCTCAGGGGGCGCAAAATGGCAAGCGTCGGCTCCA-----	383

```

Db      652 IleGlyArgTyrProArgArgAlaArgAlaAlaProArgArgAlaProAlaAlaArgAsp 671
Qy      382 CGACCTGGACATTCTTCCAGCCCT-----TGGCCTTCGCCTTCT 344
Db      672 ArgArgSerCysAlaValAlaProAlaThrAlaGluArgAlaIleTyrProGlyAsnAla 691
Qy      343 TCTTGGCCACTCGCACAGCGAGTGGCACAGGTGCA---CCAGTAGATGGACTTGA 287
Db      692 ProArgArgProGluSerArgProAlaGluGlyArgSerProArgProLeuAlaLeuGly 711
Qy      286 TCGCCAGGTGATGTAATCAGCCATCATATCGACATTCTCCCGAGTCCACCCAGGT 227
Db      712 HisProGlySerCys-----HisPro--- 718
Qy      226 CAACCCAGATGAGGTTTCGAGCGTTCGGCCAGCGGGCAGCAAGGCGCGCTTGGGGCC 167
Db      719 -----AlaArgAspArgGlyGlnCysSerGlnMetProValArg 731
Qy      166 CGTAGAGCTTCCAGCGGAGCAGCGTGAT 137
Db      732 AlaArgGluSerGlyArgAspAsnValAsp 741

```

Search completed: March 14, 2006, 01:44:31
Job time : 291 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 13, 2006, 21:32:22 ; Search time 7964 Seconds
(without alignments)
11438.278 Million cell updates/sec

Title: US-10-620-914-44

Perfect score: 1947
Sequence: 1 atggggctcggctcgtgacgg.....gogccaagaagacaactaa 1947

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_hic:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	725	37.2	743	5	BU651274 1112092D1
2	556	28.6	556	1	AV627635 AV627635
3	470.4	24.2	475	1	AV644266 AV644266
4	463.8	23.8	467	1	AV629528 AV629528
5	414	21.3	532	1	AV628087 AV628087
6	391	20.1	391	1	AV626756 AV626756
7	379	19.5	379	3	BP086130 BP086130
8	354	18.2	354	1	AV631627 AV631627
9	332	18.1	352	3	BP093875 BP093875
10	336	17.3	454	3	BP093208 BP093208
11	319	16.4	508	1	AV641303 AV641303
12	310	15.9	511	1	AV626139 AV626139
13	287	14.7	478	1	AV642589 AV642589
14	286	14.7	480	1	AV643824 AV643824
15	267	13.7	466	1	AV628989 AV628989
16	262	13.5	262	1	AV630897 AV630897
17	255	13.1	461	5	BQ824076 BQ824076
18	250	12.8	447	1	AV642307 AV642307
19	238	12.2	250	3	BP089550 BP089550
20	189	9.7	398	3	BP093467 BP093467
21	188.8	9.7	287	1	AV640882 AV640882
22	188.8	9.7	362	3	BP096956 BP096956

23	188.8	9.7	366	3	BP098395
24	188.8	9.7	377	3	BP097104
25	188.8	9.7	388	3	BP095133
26	174.4	9.0	416	6	CF556716
27	172.8	8.9	176	1	AV628222
28	143	7.3	585	3	BI721528
29	107.6	5.5	973	11	CNS07804
30	99.2	5.1	433	1	AW561229
31	88	4.5	838	11	CNS076PM
32	85.6	4.4	1060	10	CW929808
33	82	4.2	850	11	CNS06VGM
34	73.6	3.8	829	8	CV772865
35	71	3.6	488	1	AV624788
36	69.8	3.6	1895	11	DQ035307
37	69.2	3.6	841	3	BI949299
38	67.4	3.5	963	11	CNS071FB
39	67	3.4	301	2	BF072578
40	67	3.4	417	10	CZ775539
41	66	3.4	688	10	CZ735917
42	65.6	3.4	792	7	CO894488
43	64.8	3.3	528	2	BG464204
44	64.8	3.3	548	1	AW678218
45	64.8	3.3	553	2	BG240360

ALIGNMENTS

RESULT 1
BU651274

LOCUS

DEFINITION

1112092D1.y1 C. reinhardtii CC-1690 (mt+), Gamete

(normalized), Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA

sequence.

ACCESSION

BU651274

VERSION

EST.

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

BU651274 743 bp mRNA linear EST 30-SEP-2002
1112092D1.y1 C. reinhardtii CC-1690 (mt+), CC-1691 (mt-), Gamete
(normalized), Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA
sequence.

ACCESSION BU651274 GI:23363454

VERSION EST.

KEYWORDS Chlamydomonas reinhardtii

SOURCE Chlamydomonas reinhardtii

ORGANISM Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;

Chlamydomonadales; Chlamydomonadales; Chlamydomonas.

REFERENCE 1 (bases 1 to 743)

AUTHORS Grosman,A., Chang,C.-W., Davies,J., Harris,E., Hauser,C.,

Lefebvre,P., McDermott,J.P., Shrager,J., Silflow,C. and Stern,D.

TITLE Analyses of the Chlamydomonas reinhardtii Genome: A Model,

Unicellular System for Analyzing Gene Function and Regulation in

Vascular Plants. Project: 1112

JOURNAL Unpublished (2002)

COMMENT Contact: Charles Hauser

DCMB Box 91000

Duke University

Durham, NC 27708-1000

Tel: 919 613 8159

Fax: 919 613 8177

Email: chauser@duke.edu.

Location/Qualifiers

1..743

/organism="Chlamydomonas reinhardtii"

/mol_type="mRNA"

/strain="21gr (CC-1690 wild type mt+)" & 6145c (CC-1691

wild type mt-)"

/db_xref="taxon:3055"

/clone_lib="C. reinhardtii CC-1690 (mt+), CC-1691 (mt-),

Gamete (normalized), Lambda Zap II"

/note="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2:

XhoI; Gamete library was constructed by Hui Zhao, Min Lu,

Jeffrey McDermott, William J. Snell and John Davies.

Strain 21gr cells (CC-1690; mating type plus) and strain

6145c cells(CC-1691; mating type minus) that had been

growing on a light-dark cycle (13:11 L/D) in R-medium

(Sager and Granick) were separately transferred into

nitrogen-free medium at 8 hours into the light period.

PolyA mRNA was purified from each sample every 2 hours for

the next 18 hours. The mRNA was pooled and used for cDNA synthesis. The cDNA was directionally cloned into lambda Zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites. pBluescript II SK- plasmids were excised from the lambda ZAP clones by superinfection with ExAssist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al., (1996) Genome Research 6: 791-806."

ORIGIN

```

Query Match      37.2%; Score 725; DB 5; Length 743;
Best Local Similarity 99.7%; Pred. No. 6.5e-139;
Matches 736; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 681 GGAGCAGAACACCCAGGGTTCGATCCCTAGTCCGCTGCTGCGCGCCCCCTACTACGT 740
DB 7 GGAGCAGAACACCCAGGGTTCGATCCCTAGTCCGCTGCTGCGCGCCCCCTACTACGT 66
QY 741 GTGATTGGCGCGCTGCCAGCGTTGGCCACGCGCTGCACGAGGAGCGCTGGAGCGGCC 800
DB 67 GTGATTGGCGCGCTGCCAGCGTTGGCCACGCGCTGCACGAGGAGCGCTGGAGCGGCC 126
QY 801 GCCCATGTTCCCGCCCACTTCTGTACACGAGTCTGGAGGACCCCGAGCGCGATAT 860
DB 127 GCCCATGTTCCCGCCCACTTCTGTACACGAGTCTGGAGGACCCCGAGCGCGATAT 186
QY 861 GGAGGTGATGAGATCAACCCCAAGACAGGTGCTGACCTGACTAGCGCGCTGCCAA 920
DB 187 GGAGGTGATGAGATCAACCCCAAGACAGGTGCTGACCTGACTAGCGCGCGCTGCCAA 246
QY 921 TGCCCTCAACTGCTGTCAGGGGCGCGCCAGTGGTGTGCTGGTGAACCCCGC 980
DB 247 TGCCCTCAACTGCTGTCAGGGGCGCGCCAGTGGTGTGCTGGTGAACCCCGC 306
QY 981 GCAGTCGCGCTTCTGAGCTGAAGAAGTGGCCATTCAGCAGCTGAGTTGAGGACGT 1040
DB 307 GCAGTCGCGCTTCTGAGCTGAAGAAGTGGCCATTCAGCAGCTGAGTTGAGGACGT 366
QY 1041 GTGCGACCTGTTGGCGAGGGCGTGCACCGCGCATTTGAGGAGCTGTACGAGAAGCT 1100
DB 367 GTGCGACCTGTTGGCGAGGGCGTGCACCGCGCATTTGAGGAGCTGTACGAGAAGCT 426
QY 1101 GCGCGCTTCTGTCGCAACACGACCACTTCTGTCGCAAGCGCTCTGCTGACTTCCA 1160
DB 427 GCGCGCTTCTGTCGCAACACGACCACTTCTGTCGCAAGCGCTCTGCTGACTTCCA 486
QY 1161 GCAGCGCTGTACTACAGGGCGGTCATGGGCAAGCTGTGCTGGTGTGCTGCTGGC 1220
DB 487 GCAGCGCTGTACTACAGGGCGGTCATGGGCAAGCTGTGCTGGTGTGCTGCTGGC 546
QY 1221 CGTGGTCTGGGACTGGGCAAGCCGTCAAGCGCTCGCCACCGCGCCCAATGGAGGA 1280
DB 547 CGTGGTCTGGGACTGGGCAAGCCGTCAAGCGCTCGCCACCGCGCCCAATGGAGGA 606
QY 1281 GCAGCGCTGTGTCGACAGCAACATGCTCATCTTCTGTAAGAACGGGCGCCAGCC 1340
DB 607 GCAGCGCTGTGTCGACAGCAACATGCTCATCTTCTGTAAGAACGGGCGCCAGCC 666
QY 1341 GCTGGTGTGGCTGTTCGTAAGTTGTCGAGCTGGTGTCTTCAACAGGCGGTGCTGTG 1400
DB 667 GCTGGTGTGGCTGTTCGTAAGTTGTCGAGCTGGTGTCTTCAAC-AGGCGGTGCTGTG 725
QY 1401 GTTCGGCGGGCGGTGCC 1418
DB 726 GTTCNGCGGGCGGTGCC 743

```

RESULT 2

AV627635 556 bp mRNA linear EST 15-DEC-2000
 LOCUS AV627635 Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas
 DEFINITION reinhardtii cDNA clone LCL029g11_r 5', mRNA sequence.
 ACCESSION AV627635

VERSION
 KEYWORDS
 SOURCE
 ORGANISM

AV627635.1 GI:10790269

EST.

Chlamydomonas reinhardtii

Chlamydomonas reinhardtii
 Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;
 Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.

REFERENCE

1 (bases 1 to 556)

Akanuma, E., Miura, K., Kucho, K., Inoue, Y., Fukuzawa, H., Ohyama, K.,

Nakamura, Y. and Tabata, S.

Adaptation of expressed sequence tags from low-CO2 and high-CO2

generated cells of Chlamydomonas reinhardtii

DNA Res. 7 (5), 305-307 (2000)

COMMENT

Contact: Erika Asamizu

The First Laboratory for Plant Gene Research

Kazusa DNA Research Institute

Yana 1532-3, Kisarazu, Chiba 292-0812, Japan

Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES

source

1..556

/organism="Chlamydomonas reinhardtii"

/mol_type="mRNA"

/strain="C9"

/db_xref="taxon:3055"

/clone_lib="LCL029g11_r"

/note="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2:

XhoI; The cDNA library was constructed from cells cultured

in a carbon stress acclimated condition in which carbon

dioxide concentration in the bubbling gas was changed from

5% to 0.04%"

ORIGIN

```

Query Match      28.6%; Score 556; DB 1; Length 556;
Best Local Similarity 100.0%; Pred. No. 4.5e-104;
Matches 556; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 284 CGAAGTTCAAGTCCATCTAGCTGTCGACCTGTCGCACTCGCTGTGGAGGTGGCCAAAGA 343
DB 1 CGAAGTTCAAGTCCATCTAGCTGTCGACCTGTCGCACTCGCTGTGGAGGTGGCCAAAGA 60
QY 344 AGAAGCGCAGGCGCAAGGGCTGGAAGAATGTCAGGTGCTGGAGGCGCAGCTTGGCAAT 403
DB 61 AGAAGCGCAGGCGCAAGGGCTGGAAGAATGTCAGGTGCTGGAGGCGCAGCTTGGCAAT 120
QY 404 TTGCGGCCCTGAGGGCACCAGCGCTCATCACTTCTCTACTCTGCTCAGCATGATTC 463
DB 121 TTGCGGCCCTGAGGGCACCAGCGCTCATCACTTCTCTACTCTGCTCAGCATGATTC 180
QY 464 CACCGTTCCACAACGTCATCGACCGCTTGTCTGTAACCTGTCCTCAAGAGCGGCTGGTG 523
DB 181 CACCGTTCCACAACGTCATCGACCGCTTGTCTGTAACCTGTCCTCAAGAGCGGCTGGTG 240
QY 524 GCGTTGCGCATTTCTACGTGAGCGGCAAGTACGACCTGCGCCCTGCGCCAGATGCCCTGGT 583
DB 241 GCGTTGCGCATTTCTACGTGAGCGGCAAGTACGACCTGCGCCCTGCGCCAGATGCCCTGGT 300
QY 584 CGGCGCGTTTCTTCTGGCGATCGATCTTCGACATCGACACATTCATCGGCCCCCGAGC 643
DB 301 CGGCGCGTTTCTTCTGGCGATCGATCTTCGACATCGACACATTCATCGGCCCCCGAGC 360
QY 644 GCGCGCGCTACCTGGAGCAGAAAGCTGGAGCGCGTGTGGGAGCAGAAACACCCAGGGTTCGA 703
DB 361 GCGCGCGCTACCTGGAGCAGAAAGCTGGAGCGCGTGTGGGAGCAGAAACACCCAGGGTTCGA 420
QY 704 TCCCTACGTGCGCTGCTGCGGCCCTTACTACGTGTGGAATTGGCGCGCTGCGCCAGCG 763
DB 421 TCCCTACGTGCGCTGCTGCGGCCCTTACTACGTGTGGAATTGGCGCGCTGCGCCAGCG 480
QY 764 TTGGCCACGCGCTTCAGGAGGCGGTGGAGCGGCGCCGCTTCCCGCCCACTTCC 823
DB 481 TTGGCCACGCGCTTCAGGAGGCGGTGGAGCGGCGCCGCTTCCCGCCCACTTCC 540

```

Db	362	TC	CCCTACGTGCGTGGCTGGCGCCCTTACTACGTGTGATGTCNCGCTGCCCGCCACGC	421
Qy	764	TT	GGCCACGCGCTTGCACACGAGCGGTGAGGGCGCGCCCATGTTCCCGCCCA	817
Db	422	TT	GGCCACGCGCTTGCACACGAGCGGTGAGGGCGCGCCCATGTTCCCGCCCA	475
RESULT 4				
LOCUS	AV629528			
DEFINITION	AV629528 Chlamydomonas reinhardtii 5' to 0.04% CO2 Chlamydomonas reinhardtii cDNA clone LCL060c10_r 5', mRNA sequence.			
ACCESSION	AV629528			
VERSION	1			
KEYWORDS	EST.			
SOURCE	AV629528.1	GI:10792162		
ORGANISM	Chlamydomonas reinhardtii Chlamydomonas reinhardtii Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas. 1 (bases 1 to 467) Asamizu, Y., Miura, K., Kucho, K., Inoue, Y., Fukuzawa, H., Ohyama, K., Nakamura, Y. and Tabata, S. Generation of expressed sequence tags from low-CO2 and high-CO2 adapted cells of Chlamydomonas reinhardtii DNA Res. 7 (5), 305-307 (2000)			
REFERENCE	Contact: Erika Asamizu The First Laboratory for Plant Gene Research Kazusa DNA Research Institute Yana 1532-3, Kisarazu, Chiba 292-0812, Japan Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/			
AUTHORS	Location/Qualifiers 1. .467 /organism="Chlamydomonas reinhardtii" /mol_type="mRNA" /strain="C9" /db_xref="taxon:3055" /clone="LCL060c10_r" /note="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2: XhoI; The cDNA library was constructed from cells cultured in a carbon stress acclimatized condition in which carbon dioxide concentration in the bubbling gas was changed from 5% to 0.04%"			
TITLE	1. .467			
JOURNAL	1. .467			
PUBMED	11089912			
COMMENT	Contact: Erika Asamizu The First Laboratory for Plant Gene Research Kazusa DNA Research Institute Yana 1532-3, Kisarazu, Chiba 292-0812, Japan Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/			
FEATURES	Location/Qualifiers 1. .467 /organism="Chlamydomonas reinhardtii" /mol_type="mRNA" /strain="C9" /db_xref="taxon:3055" /clone="LCL060c10_r" /note="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2: XhoI; The cDNA library was constructed from cells cultured in a carbon stress acclimatized condition in which carbon dioxide concentration in the bubbling gas was changed from 5% to 0.04%"			
SOURCE	1. .467			
ORIGIN	1. .467			
Query Match	23.8%;	Score 463.8;	DB 1;	Length 467;
Best Local Similarity	99.8%;	Pred. No. 4.5e-85;		
Matches 465;	Conservative	0;	Mismatches 2;	Indels 0; Gaps 0;
Qy	178	GCTGCGCTTTGCTGCCCGCTGGCGGAGCGCTCGAACCTCATCTGGGTGACCTGGGTGGT	237	
Db	1	GCTGCGAGTTGCTGCCCGCTGGCGGAGCGCTCGAACCTCATCTGGGTGACCTGGGTGGT	60	
Qy	238	GGCACTGGGAGAAATGTCGATATGATGGCTGATTATCATCGACCTGGCGGAAGTTCAAGTCC	297	
Db	61	GGCACTGGGAGAAATGTCGATATGATGGCTGATTATCATCGACCTGGCGGAAGTTCAAGTCC	120	
Qy	298	ATTCAGTGGTGCACCTGTGCGCACTGCTGTGCGAGGTGGCCGAAGAGAGGCGGAAGGCC	357	
Db	121	ATTCAGTGGTGCACCTGTGCGCACTGCTGTGCGAGGTGGCCGAAGAGAGGCGGAAGGCC	180	
Qy	358	AAGGGCTGGAAGAAATGTCAGAGTCTGAGAGCGCGCTTGGCCAAATTTGCGCCCCCTGAG	417	
Db	181	AAGGGCTGGAAGAAATGTCAGAGTCTGAGAGCGCGCTTGGCCAAATTTGCGCCCCCTGAG	240	
Qy	418	GGCACCGCAGCGCTCATCACTTCTCTCTACTCTGCTCACGATGATTCACCGTTTCCAAAC	477	
Db	241	GGCACCGCAGCGCTCATCACTTCTCTCTACTCTGCTCACGATGATTCACCGTTTCCAAAC	300	
Qy	478	GTCAATCGACAGCGCTTGTGTAATGTCGTGCCAAGACGGCTGTGGGGCGTTGCCGACTTC	537	
Db	301	GTCAATCGACAGCGCTTGTGTAATGTCGTGCCAAGACGGCTGTGGGGCGTTGCCGACTTC	360	

[illegible]

QY 538 TAGCTGAGCGGCAAGTACGACCTGCCCTGCGCCAGATGCCCTGCTGCGCGCGTTCTTTC 597
 Db 361 TAGCTGAGCGGCAAGTACGACCTGCCCTGCGCCAGATGCCCTGCTGCGCGCGTTCTTTC 420

QY 598 TGGCGATCGATCTTCGACATCGACAACTGACATCGCGCCCGGAGCG 644
 Db 421 TGGCGATCGATCTTCGACATCGACAACTGACATCGCGCCCGGAGCG 467

RESULT 5
 AV628087 532 bp mRNA linear EST 15-DEC-2000
 LOCUS AV628087 Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas
 DEFINITION reinhardtii cDNA clone LCL036e03_r 5', mRNA sequence.
 ACCESSION AV628087
 VERSION AV628087.1 GI:10790721
 KEYWORDS EST.
 SOURCE Chlamydomonas reinhardtii
 ORGANISM Chlamydomonas reinhardtii
 Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;
 Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.
 REFERENCE 1 (bases 1 to 532)
 AUTHORS Asamizu, E., Miura, K., Kucho, K., Inoue, Y., Fukuzawa, H., Ohyama, K.,
 Nakamura, Y. and Tabata, S.
 TITLE Generation of expressed sequence tags from low-CO2 and high-CO2
 adapted cells of Chlamydomonas reinhardtii
 JOURNAL DNA Res. 7 (5), 305-307 (2000)
 PUBMED 11089912
 COMMENT Contact: Erika Asamizu
 The First Laboratory for Plant Gene Research
 Kazusa DNA Research Institute
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
 Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES
 source
 1. 532
 Location/Qualifiers
 /organism="Chlamydomonas reinhardtii"
 /mol_type="mRNA"
 /strain="C9"
 /db_xref="taxon:3055"
 /clone="LCL036e03_r"
 /notes="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2:
 XhoI; The cDNA library was constructed from cells cultured
 in a carbon stress acclimatized condition in which carbon
 dioxide concentration in the bubbling gas was changed from
 5% to 0.04%"

ORIGIN
 Query Match 21.3%; Score 414; DB 1; Length 532;
 Best Local Similarity 89.8%; Pred. No. 8.5e-75;
 Matches 478; Conservative 0; Mismatches 0; Indels 54; Gaps 1;
 7 TCGGGTCGTGACGGCGCGCTCGGAGCTACACCAAGAGAACTTCTCCCTGGAGAGCTC 66
 Db 1 TCGGGTCGTGACGGCGCGCTCGGAGCTACACCAAGAGAACTTCTCCCTGGAGAGCTC 60
 QY 67 AAGCTCAGCAGCATGAAGATGACCTGACCGTTCTGCGCCATATGTGCTGGCAGCAAG 126
 Db 61 AAGCTCAGCAGCATGAAGATGACCTGACCGTTCTGCGCCATATGTGCTGGCAGCAAG 120
 QY 127 AAGGGCATGATCAGCTGCTCGCTCGAGAGCTTCTACGGGCCCGCAGCGCTGCTCT-- 184
 Db 121 AAGGGCATGATCAGCTGCTCGCTCGAGAGCTTCTACGGGCCCGCAGCGCTGCTCTAT 180
 QY 185 -----TTGCTGCC 192
 Db 181 GATGCTTTCCGGTCGCGGTTCTCTGCGGTGCGAGGCCCATGCTCGCTGCAAGTTGCTGCC 240
 QY 193 CGCCTGCGCGAGCGCTCGAACCTCATCTGGGTTGACCTGGGTGCTGSCACTGGGGAGAT 252
 Db 241 CGCCTGCGCGAGCGCTCGAACCTCATCTGGGTTGACCTGGGTGCTGSCACTGGGGAGAT 300

QY 253 GTCGATATGATGGCTGATTACATCGACTGGCGAAGTTCAAGTCCATCTACGTGGTCGAC 312
 Db 301 GTCGATATGATGGCTGATTACATCGACTGGCGAAGTTCAAGTCCATCTACGTGGTCGAC 360

QY 313 CTGTGCCACTCGTGTGCGAGGTGGCCAAAGAGAAGCGCAAGCGCTGGAGAGAT 372
 Db 361 CTGTGCCACTCGTGTGCGAGGTGGCCAAAGAGAAGCGCAAGCGCTGGAGAGAT 420

QY 373 GTCCAGGTGCTGGAGGCCCGACGCTTGCCTCAATTGCGCCCTTGAGGGCACCGGACGCTC 432
 Db 421 GTCCAGGTGCTGGAGGCCCGACGCTTGCCTCAATTGCGCCCTTGAGGGCACCGGACGCTC 480

QY 433 ATCACCTTCTCTACTCTCGTCTCAGATGATTCCACCGTTCCACAACGTCATCG 484
 Db 481 ATCACCTTCTCTACTCTCGTCTCAGATGATTCCACCGTTCCACAACGTCATCG 532

RESULT 6
 AV626756 391 bp mRNA linear EST 15-DEC-2000
 LOCUS AV626756 Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas
 DEFINITION reinhardtii cDNA clone LCL014h09_r 5', mRNA sequence.
 ACCESSION AV626756
 VERSION AV626756.1 GI:10789036
 KEYWORDS EST.
 SOURCE Chlamydomonas reinhardtii
 ORGANISM Chlamydomonas reinhardtii
 Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;
 Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.
 REFERENCE 1 (bases 1 to 391)
 AUTHORS Asamizu, E., Miura, K., Kucho, K., Inoue, Y., Fukuzawa, H., Ohyama, K.,
 Nakamura, Y. and Tabata, S.
 TITLE Generation of expressed sequence tags from low-CO2 and high-CO2
 adapted cells of Chlamydomonas reinhardtii
 JOURNAL DNA Res. 7 (5), 305-307 (2000)
 PUBMED 11089912
 COMMENT Contact: Erika Asamizu
 The First Laboratory for Plant Gene Research
 Kazusa DNA Research Institute
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
 Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES
 source
 1. 391
 Location/Qualifiers
 /organism="Chlamydomonas reinhardtii"
 /mol_type="mRNA"
 /strain="C9"
 /db_xref="taxon:3055"
 /clone="LCL014h09_r"
 /notes="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2:
 XhoI; The cDNA library was constructed from cells cultured
 in a carbon stress acclimatized condition in which carbon
 dioxide concentration in the bubbling gas was changed from
 5% to 0.04%"

ORIGIN
 Query Match 20.1%; Score 391; DB 1; Length 391;
 Best Local Similarity 100.0%; Pred. No. 4.6e-70;
 Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 229 CTGGGTGGTGGCACTGGGGAGAAATGCGATATATGATGCTGATTCATCGACCTGGCGAAG 288
 Db 1 CTGGGTGGTGGCACTGGGGAGAAATGCGATATATGATGCTGATTCATCGACCTGGCGAAG 60
 QY 289 TTCAAGTCCATCTACGTGCTGCGACCTGTCACCTGCGCTGCGAGGTGGCCAGAGAAG 348
 Db 61 TTCAAGTCCATCTACGTGCTGCGACCTGTCGATATGATGCTGATTCATCGACCTGGCGAAG 120
 QY 349 GCGAAGGCCCAAGCGCTGGAAGAATGTCCAGGTGCTGGAGGCCGACGCTTGCCTATTTCGCG 408
 Db 121 GCGAAGGCCCAAGCGCTGGAAGAATGTCCAGGTGCTGGAGGCCGACGCTTGCCTATTTCGCG 180
 QY 409 CCCCCTGAGGGGCAACCGCGACGCTCATCACCTTCTCTACTCGCTCAGATGATTCCACCG 468

```

181 181 CCCCCCTGAGGCGACCGCGCTCATACCTTCTCTACTCGCTACGATGATTCACCG 240
QY 469 TTCCACACGCTCATCGACCGAGCTTGCTCGTACTGTCTCCCAAGCGCTGTGGGCGTT 528
Db 241 TTCCACACGCTCATCGACCGAGCTTGCTCGTACTGTCTCCCAAGCGCTGTGGGCGTT 300
QY 529 GCGGACTTCTAGCTGAGCGGCAAGTAGACGCTGCGGCGGCGGCGAGATGCGGCGCG 588
Db 301 GCGGACTTCTAGCTGAGCGGCAAGTAGACGCTGCGGCGGCGGCGAGATGCGGCGCG 360
QY 589 CGTTTCTTCTGCGGATCGATCTTCGACATCG 619
Db 361 CGTTTCTTCTGCGGATCGATCTTCGACATCG 391

```

```

RESULT 7
BP086130 379 bp mRNA linear EST 30-JUN-2004
LOCUS BP086130 Chlamydomonas reinhardtii C9 various conditions
DEFINITION Chlamydomonas reinhardtii cDNA clone MX002e03_r 5', mRNA sequence.
ACCESSION BP086130
VERSION BP086130.1 GI:49458217
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii
ORGANISM Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;
Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.
REFERENCE 1 (bases 1 to 379)
AUTHORS Asanizu, E., Nakamura, Y., Miura, K., Fukuzawa, H., Fujiwara, S.,
Hirono, M., Iwamoto, K., Matsuda, Y., Minagawa, J., Shimogawara, K.,
Takahashi, Y., and Tabata, S.
TITLE Establishment of Publicly Available cDNA Material and Information
Resource of Chlamydomonas reinhardtii (Chlorophyta), to Facilitate
Gene Function Analysis
JOURNAL Phycologia (2004) In press
COMMENT Contact: Erika Asanizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asanizu@kazusa.or.jp. URL: http://www.kazusa.or.jp/en/plant/.
FEATURES
source
1. 379
/organism="Chlamydomonas reinhardtii"
/mol_type="mRNA"
/strain="C9"
/db_xref="taxon:3055"
/clone="MX002e03_r"
/clone_lib="Chlamydomonas reinhardtii C9 various
conditions"
/note="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2:
XhoI; The cDNA library was made from a mixture of cells
grown under various conditions"

```

```

ORIGIN
Query Match 19.5%; Score 379; DB 3; Length 379;
Best Local Similarity 100.0%; Pred. No. 1.4e-67;
Matches 379; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1449 CCGCATCCCATGAGAACTACATCGCGCGCACCATGGAGCGGCGGCGAGACTCGCA 1508
Db 1 CCGCATCCCATGAGAACTACATCGCGCGCACCATGGAGCGGCGGCGAGACTCGCA 60
QY 1509 CGTGGCGAAGCAGAACTACTTCTACTACAACTGCCTCACCGGCAAGTCTCTCGCGACAA 1568
Db 61 CGTGGCGAAGCAGAACTACTTCTACTACAACTGCCTCACCGGCAAGTCTCTCGCGACAA 120
QY 1569 CTGCCCCACCTACTCGCGGAGCGGCGCTTGCCACCCCTCAAGAGTGGCGTGGTGAACAA 1628
Db 121 CTGCCCCACCTACTCGCGGAGCGGCGCTTGCCACCCCTCAAGAGTGGCGTGGTGAACAA 180
QY 1629 CTTGACCCGTCTCCACCAACTTCTTCTATGAGGAGCTCAAGCGCGCACCTACCAAGGT 1688

```

```

Db 181 CTTGACCCGTCTCCACCAACTTCTTCTATGAGGAGCTCAAGCGCGCACCTACCAAGGT 240
QY 1689 GATTTCTGATGGACACGCTGAGCTGGCTGGATATGCCGTGGCCCAAGAGTGGCCGAGTG 1748
Db 241 GATTTCTGATGGACACGCTGAGCTGGCTGGATATGCCGTGGCCCAAGAGTGGCCGAGTG 300
QY 1749 CTTGGCCAAAGCAGGTTGCGCGGCGGCGCATGCTCATCTGGCGCTCGCCTCCCTCAGCCC 1808
Db 301 CTTGGCCAAAGCAGGTTGCGCGGCGGCGCATGCTCATCTGGCGCTCGCCTCCCTCAGCCC 360
QY 1809 GCCCTACGCGGAGCTGATC 1827
Db 361 GCCCTACGCGGAGCTGATC 379

```

```

RESULT 8
AV631627 354 bp mRNA linear EST 15-DEC-2000
LOCUS AV631627 Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas
DEFINITION reinhardtii cDNA clone LCL097c01_r 5', mRNA sequence.
ACCESSION AV631627
VERSION AV631627.1 GI:10794261
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii
ORGANISM Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;
Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.
REFERENCE 1 (bases 1 to 354)
AUTHORS Asanizu, E., Miura, K., Kucho, K., Inoue, Y., Fukuzawa, H., Ohshima, K.,
Nakamura, Y., and Tabata, S.
TITLE Generation of expressed sequence tags from low-CO2 and high-CO2
adapted cells of Chlamydomonas reinhardtii
JOURNAL DNA Res. 7 (5), 305-307 (2000)
COMMENT Contact: Erika Asanizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asanizu@kazusa.or.jp. URL: http://www.kazusa.or.jp/en/plant/.
FEATURES
source
1. 354
/organism="Chlamydomonas reinhardtii"
/mol_type="mRNA"
/strain="C9"
/db_xref="taxon:3055"
/clone="LCL097c01_r"
/clone_lib="Chlamydomonas reinhardtii 5% to 0.04% CO2"
/note="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2:
XhoI; The cDNA library was constructed from cells cultured
in a carbon stress acclimatized condition in which carbon
dioxide concentration in the bubbling gas was changed from
5% to 0.04%"

```

```

ORIGIN
Query Match 18.2%; Score 354; DB 1; Length 354;
Best Local Similarity 100.0%; Pred. No. 1.9e-62;
Matches 354; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 329 GCGAGTGGCCAAAGAGGAGGCGCAAGGCGCTGGAGGATGTCACAGGTGCTGGAGG 388
Db 1 GCGAGTGGCCAAAGAGGAGGCGCAAGGCGCTGGAGGATGTCACAGGTGCTGGAGG 60
QY 389 CCGAGCTTGGCCAAATTTGCGCCCTCGGCGGCGACCGCGCATCATCTCTCTACT 448
Db 61 CCGAGCTTGGCCAAATTTGCGCCCTCGGCGGCGACCGCGCATCATCTCTCTACT 120
QY 449 CGCTCAGCATGATTCACCGTTCCCAAGCTCATGACCAAGGCTTCTCGTACTCTGCC 508
Db 121 CGCTCAGCATGATTCACCGTTCCCAAGCTCATGACCAAGGCTTCTCGTACTCTGCC 180
QY 509 AAGAGCGGCTGGTGGCGGTTGCGGACTTCTAGTCAGCGGCGCAAGTACGACCTGCGCCCTGC 568
Db 181 AAGAGCGGCTGGTGGCGGTTGCGGACTTCTAGTCAGCGGCGCAAGTACGACCTGCGCCCTGC 240

```

```

QY 569 GCAGATGCCCTGCTGCGCCGCTTTCTTCTGGCGATCGATCTTCGACATCGACAACATTG 628
Db 241 GCAGATGCCCTGCTGCGCCGCTTTCTTCTGGCGATCGATCTTCGACATCGACAACATTG 300

QY 629 ACATCGCCCCCGAGCGCGCGCTTACCTGGAGCAGAGCTGGAGCGCGTGTGGG 682
Db 301 ACATCGCCCCCGAGCGCGCGCTTACCTGGAGCAGAGCTGGAGCGCGTGTGGG 354

RESULT 9
BP093875 352 bp mRNA linear EST 30-JUN-2004
LOCUS BP093875 Chlamydomonas reinhardtii C9 various conditions
DEFINITION Chlamydomonas reinhardtii cDNA clone MXL015a05_r 5', mRNA sequence.
ACCESSION BP093875
VERSION BP093875.1
KEYWORDS GI:49465962
SOURCE Chlamydomonas reinhardtii
ORGANISM Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;
Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.
REFERENCE 1 (bases 1 to 352)
AUTHORS Asamizu, E., Iwamoto, K., Miura, K., Fukuzawa, H., Fujiwara, S.,
Takahashi, Y. and Tabata, S.
Hirono, M., Minagawa, J., Shimogawara, K.,
Establishment of Publicly Available cDNA Material and Information
Resource of Chlamydomonas reinhardtii (Chlorophyta), to Facilitate
Gene Function Analysis
Phycologia (2004) In press
Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES
source
Location/Qualifiers
1..352
/organism="Chlamydomonas reinhardtii"
/mol_type="mRNA"
/strain="C9"
/db_xref="taxon:3055"
/clone="MXL015a05_r"
/clone_lib="Chlamydomonas reinhardtii C9 various
conditions"
/note="vector: pBluescriptII SK-; Site 1: EcoRI; Site 2:
XhoI; The cDNA library was made from a mixture of cells
grown under various conditions"

ORIGIN
Query Match 18.1%; Score 352; DB 3; Length 352;
Best Local Similarity 100.0%; Pred. No. 5e-62;
Matches 352; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1257 CGCCAAACGCGCCCAATGGAGGAGCAGCGCGCTGTGGACAGCAACATGCTCATCCA 1316
Db 1 CGCCAAACGCGCCCAATGGAGGAGCAGCGCGCTGTGGACAGCAACATGCTCATCCA 60

QY 1317 CTTGCTGAAGAACGGGCGCCAGCGCTGGTGTGGCTGTCTGTCAGTTCTGTGACCTGGT 1376
Db 61 CTTGCTGAAGAACGGGCGCCAGCGCGCTGGTGTGGCTGTCTGTCAGTTCTGTGACCTGGT 120

QY 1377 GCTCTTCAACAAAGCGCTGCTGGTTTCGGCGCGCGCTGCGGGCAAGCAGTACGCGCT 1436
Db 121 GCTCTTCAACAAAGCGCTGCTGGTTTCGGCGCGCGCTGCGGGCAAGCAGTACGCGCT 180

QY 1437 GATCAAGCGGACGGGCATCCCATTTGAGAACTACATCGCGCGCAACATGACGCGTGGC 1496
Db 181 GATCAAGCGGACGGGCATCCCATTTGAGAACTACATCGCGCGCAACATGACGCGTGGC 240

QY 1497 GGAGAACTCGCACTGGCGCAGCAAGTACTTCTTACTAGAACTGCTTCAACCTCAGCGCAAGTT 1556
Db 241 GGAGAACTCGCACTGGCGCAGCAAGTACTTCTTACTAGAACTGCTTCAACCTCAGCGCAAGTT 300

```

```

QY 1557 CCTGCGCGCAAACTGCCCACTTACCTGCGCGAGCGCGCTTGGCCACCTC 1608
Db 301 CCTGCGCGCAAACTGCCCACTTACCTGCGCGAGCGCGCTTGGCCACCTC 352

RESULT 10
BP093208 454 bp mRNA linear EST 30-JUN-2004
LOCUS BP093208 Chlamydomonas reinhardtii C9 various conditions
DEFINITION Chlamydomonas reinhardtii cDNA clone MXL005a04_r 5', mRNA sequence.
ACCESSION BP093208
VERSION BP093208.1
KEYWORDS GI:49465295
SOURCE Chlamydomonas reinhardtii
ORGANISM Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;
Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.
REFERENCE 1 (bases 1 to 454)
AUTHORS Asamizu, E., Iwamoto, K., Miura, K., Fukuzawa, H., Fujiwara, S.,
Hirono, M., Minagawa, J., Shimogawara, K.,
Takahashi, Y. and Tabata, S.
Establishment of Publicly Available cDNA Material and Information
Resource of Chlamydomonas reinhardtii (Chlorophyta), to Facilitate
Gene Function Analysis
Phycologia (2004) In press
Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES
source
Location/Qualifiers
1..454
/organism="Chlamydomonas reinhardtii"
/mol_type="mRNA"
/strain="C9"
/db_xref="taxon:3055"
/clone="MXL005a04_r"
/clone_lib="Chlamydomonas reinhardtii C9 various
conditions"
/note="vector: pBluescriptII SK-; Site 1: EcoRI; Site 2:
XhoI; The cDNA library was made from a mixture of cells
grown under various conditions"

ORIGIN
Query Match 17.3%; Score 336; DB 3; Length 454;
Best Local Similarity 86.3%; Pred. No. 1e-58;
Matches 385; Conservative 0; Mismatches 52; Indels 9; Gaps 1;

QY 69 GCTCAGCAGCATCAAGATGACCTGACCGTTCTGCGCCATATGTGTTGGCGCAAGAA 128
Db 9 GTTTCGGCAGCAAGAGGGCGATGATCACGCTGCTCGCTGGAGAGCTTCTACGGGCCCA 68

QY 129 GGGCGATGATCAGCTGCTCGCTGGAGAGCTTC-----TACGGGCCCAAGCGCGC 179
Db 69 GGGCGCTGCTATGATGCTTTCCGGTGTGCGGTCTCTCTGGGTGCGAGGCCATGCTCGC 128

QY 180 TGCCTTTGCTGCCCGCTGCGCGAGCGCTCGAACCTCATCTGGGTTCACCTGGGTGGTGG 239
Db 129 TGCAGTTGCTGCCCGCTGCGCGAGCGCTCGAACCTCATCTGGGTTCACCTGGGTGGTGG 188

QY 240 CACTGGGGAGAAATGTGCATATGATGCTGATATTCATCGACTGGCGAAGTTCAAGTCCAT 299
Db 189 CACTGGGGAGAAATGTGCATATGATGCTGATATTCATCGACTGGCGAAGTTCAAGTCCAT 248

QY 300 CTAGTGGTGCACCTGTGCCACTGCTGCGAGTGGCCCAAGAGAGCGAAGGCCAA 359
Db 249 CTAGTGGTGCACCTGTGCCACTGCTGCGAGTGGCCCAAGAGAGCGAAGGCCAA 308

QY 360 GGGCTGGAAGAAATGTCCAGGTTCGTGGAGGCGCGAGCTTGCCAAATTTCGCCCTGAGGG 419
Db 309 GGGCTGGAAGAAATGTCCAGGTTCGTGGAGGCGCGAGCTTGCCAAATTTCGCCCTGAGGG 368

QY 420 CACGCGAGCGCTCATACCTTTCTCTACTCGCTCAGCATGATTCACCGTTCACAACGT 479

```

```

ORIGIN
Query Match      15.9%; Score 310; DB 1; Length 511;
Best Local Similarity 87.4%; Pred. No. 2.4e-53;
Matches 374; Conservative 0; Mismatches 0; Indels 54; Gaps 1;

QY      1  ATGGGGTCGGGTCGTGACGGCGGGCTCGAGCTACACCAAGAAGAACTTCTCCCTGGAG 60
      |||||
Db      84  ATGGGGTCGGGTCGTGACGGCGGGCTCGAGCTACACCAAGAAGAACTTCTCCCTGGAG 143

QY      61  AAGCTCAAGCTCAGCAGCATGAAGATGACCTGACCGTTCTCGGCCATATGTGGTTCGGC 120
      |||||
Db     144  AAGCTCAAGCTCAGCAGCATGAAGATGACCTGACCGTTCTCGGCCATATGTGGTTCGGC 203

QY     121  ACGAAGAGGGCGGATGATCACCTGCTCGCTGGAGAGCTTCTACGGGCCCCAGGCCGCT 180
      |||||
Db     204  ACGAAGAGGGCGGATGATCACCTGCTCGCTGGAGAGCTTCTACGGGCCCCAGGCCGCT 263

QY     181  GCCT-----TT 186
      |||||
Db     264  GCCTATGATCTTTCGGGTCGGGTTCTCTGGGGTCGAGGCCCATGCTCGCTCGAGTT 323

QY     187  GTCGCCCGCTGGCGAGCGCTCGAAGCTCATCTCGGTTGACCTGGGTGGTGCACCTGGG 246
      |||||
Db     324  GTCGCCCGCTGGCGAGCGCTCGAAGCTCATCTCGGTTGACCTGGGTGGTGCACCTGGG 383

```

```

Qy 247 GAGAAATGTCATATGATGCTGATTACATGACCTGGCGAAGTTCAAGTCCATCTACGTG 306
Db 384 GAGAAATGTCATATGATGCTGATTACATGACCTGGCGAAGTTCAAGTCCATCTACGTG 443
Qy 307 GTCGACCTGTGCCACTCGCTGTGCGAGGTGGCCCAAGGAAGGCAAGGGCTGG 366
Db 444 GTCGACCTGTGCCACTCGCTGTGCGAGGTGGCCCAAGGAAGGCAAGGGCTGG 503
Qy 367 AAGAAATGT 374
Db 504 AAGAAATGT 511

RESULT 13
AV642589
LOCUS AV642589 478 bp mRNA linear EST 15-DEC-2000
DEFINITION AV642589 Chlamydomonas reinhardtii 5% CO2 Chlamydomonas reinhardtii
CDNA clone HCL054c11_r 5', mRNA sequence.
ACCESSION AV642589
VERSION AV642589.1 GI:10785917
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii
ORGANISM Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;
Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.
REFERENCE 1 (bases 1 to 478)
AUTHORS Nakamura,Y., Miura,K., Kucho,K., Inoue,Y., Fukuzawa,H., Ohyama,K.,
Generation of expressed sequence tags from low-CO2 and high-CO2
adapted cells of Chlamydomonas reinhardtii
JOURNAL DNA Res. 7 (5), 305-307 (2000)
PUBMED 11089912
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES
source
Location/Qualifiers
1..478
/organism="Chlamydomonas reinhardtii"
/mol_type="mRNA"
/strain="C9"
/db_xref="taxon:3055"
/clone="HCL054c11_r"
/notes="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2:
XhoI; The cDNA library was constructed from cells cultured
in a medium with bubbling air containing 5% carbon
dioxide"

ORIGIN
Query Match 14.7%; Score 287; DB 1; Length 478;
Best Local Similarity 86.7%; Pred. No. 1.3e-48;
Matches 351; Conservative 0; Mismatches 0; Indels 54; Gaps 1;

Qy 1 ATGGGGTCGGTCTGTGACGGCGCGCTGCGAGCTACACCAAGAGAAGACTTCTCCCTGGAG 60
Db 74 ATGGGGTCGGTCTGTGACGGCGCGCTGCGAGCTACACCAAGAGAAGACTTCTCCCTGGAG 133
Qy 61 AAGCTCAAGCTCAGCAGCATGAAGGATGACCTGACCGTTCTGCGCCATATGTGTTCCGC 120
Db 134 AAGCTCAAGCTCAGCAGCATGAAGGATGACCTGACCGTTCTGCGCCATATGTGTTCCGC 193
Qy 121 AGCAAGAAGGGCGATGATCAGCTGTGCGCTGAGAGCTTCTACGGGCCCGCCGCGCT 180
Db 194 AGCAAGAAGGGCGATGATCAGCTGTGCGCTGAGAGCTTCTACGGGCCCGCCGCGCT 253
Qy 181 GCCT-
Db 254 GCCTATGATGTTTCCGTCGGGTTCTCTGCGGTCGAGGCCCATGCTCGTGCAGTT 313
Qy 187 GCTGCCCGCTGGCCGAGCGCTCGAACTCATATCGGGTTGACTGGGTGGGCACTGGG 246

```

```

Db 314 GCTGCCCGCTGGCCGAGCGCTCGAACCTCATCTGGGTTGACCTGGTGGTGGCACTGGG 373
Qy 247 GAGAATGTCATATGATGCTGATTACATGACCTGGCGAAGTTCAAGTCCATCTACGTG 306
Db 374 GAGAATGTCATATGATGCTGATTACATGACCTGGCGAAGTTCAAGTCCATCTACGTG 433
Qy 307 GTCGACCTGTGCCACTCGCTGTGCGAGGTGGCCCAAGGAAGGCG 351
Db 434 GTCGACCTGTGCCACTCGCTGTGCGAGGTGGCCCAAGGAAGGCG 478

RESULT 14
AV643824
LOCUS AV643824 480 bp mRNA linear EST 15-DEC-2000
DEFINITION AV643824 Chlamydomonas reinhardtii 5% CO2 Chlamydomonas reinhardtii
CDNA clone HCL076g02_r 5', mRNA sequence.
ACCESSION AV643824
VERSION AV643824.1 GI:10787152
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii
ORGANISM Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;
Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.
REFERENCE 1 (bases 1 to 480)
AUTHORS Nakamura,Y. and Tabata,S.
Generation of expressed sequence tags from low-CO2 and high-CO2
adapted cells of Chlamydomonas reinhardtii
JOURNAL DNA Res. 7 (5), 305-307 (2000)
PUBMED 11089912
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES
source
Location/Qualifiers
1..480
/organism="Chlamydomonas reinhardtii"
/mol_type="mRNA"
/strain="C9"
/db_xref="taxon:3055"
/clone="HCL076g02_r"
/notes="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2:
XhoI; The cDNA library was constructed from cells cultured
in a medium with bubbling air containing 5% carbon
dioxide"

ORIGIN
Query Match 14.7%; Score 286; DB 1; Length 480;
Best Local Similarity 86.6%; Pred. No. 2.1e-48;
Matches 350; Conservative 0; Mismatches 0; Indels 54; Gaps 1;

Qy 1 ATGGGGTCGGTCTGTGACGGCGCGCTGCGAGCTACACCAAGAGAAGACTTCTCCCTGGAG 60
Db 77 ATGGGGTCGGTCTGTGACGGCGCGCTGCGAGCTACACCAAGAGAAGACTTCTCCCTGGAG 136
Qy 61 AAGCTCAAGCTCAGCAGCATGAAGGATGACCTGACCGTTCTGCGCCATATGTGTTCCGC 120
Db 137 AAGCTCAAGCTCAGCAGCATGAAGGATGACCTGACCGTTCTGCGCCATATGTGTTCCGC 196
Qy 121 AGCAAGAAGGGCGATGATCAGCTGTGCGCTGAGAGCTTCTACGGGCCCGCCGCGCT 180
Db 197 AGCAAGAAGGGCGATGATCAGCTGTGCGCTGAGAGCTTCTACGGGCCCGCCGCGCT 256
Qy 181 GCCT-
Db 257 GCCTATGATGTTTCCGTCGGGTTCTCTGCGGTCGAGGCCCATGCTCGTGCAGTT 316
Qy 187 GCTGCCCGCTGGCCGAGCGCTCGAACTCATCTGGGTTGACCTGGTGGGCACTGGG 246
Db 317 GCTGCCCGCTGGCCGAGCGCTCGAACTCATCTGGGTTGACCTGGTGGGCACTGGG 376

```

```
QY 247 GAGATGTCATATGATGGCTGATTACATCGACTGGCGAAGTTCAAGTCCATCTACGTG 306
Db 377 GAGAAATGTCGATATGATGGCTGATTACATCGACTGGCGAAGTTCAAGTCCATCTACGTG 436
QY 307 GTCGACCTGTGCCACTCGCTGTGCGAGTGGCCAAAGAAGGC 350
Db 437 GTCGACCTGTGCCACTCGCTGTGCGAGTGGCCAAAGAAGGC 480

RESULT 15
LOCUS AV628989 466 bp mRNA linear EST 15-DEC-2000
DEFINITION AV628989 Chlamydomonas reinhardtii 5' to 0.04% CO2 Chlamydomonas
reinhardtii cDNA clone LCL050a06_r 5', mRNA sequence.
ACCESSION AV628989
VERSION AV628989
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii
ORGANISM Chlamydomonas reinhardtii
REFERENCE 1 (bases 1 to 466)
AUTHORS Asamizu, E., Miura, K., Kucho, K., Inoue, Y., Fukuzawa, H., Ohyama, K.,
Nakamura, Y. and Tabata, S.
TITLE Generation of expressed sequence tags from low-CO2 and high-CO2
adapted cells of Chlamydomonas reinhardtii
JOURNAL DNA Res. 7 (5), 305-307 (2000)
PUBMED 11089912
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES
source
1..466
/organism="Chlamydomonas reinhardtii"
/mol_type="mRNA"
/strain="C9"
/db_xref="taxon:3055"
/clone="LCL050a06_r"
/clone_lib="Chlamydomonas reinhardtii 5' to 0.04% CO2"
/note="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2:
XhoI; The cDNA library was constructed from cells cultured
in a carbon stress acclimatized condition in which carbon
dioxide concentration in the bubbling gas was changed from
5% to 0.04%"

ORIGIN
Query Match 13.7%; Score 267; DB 1; Length 466;
Best Local Similarity 86.0%; Pred. No. 1.7e-44;
Matches 331; Conservative 0; Mismatches 0; Indels 54; Gaps 1;

QY 1 ATGGGGTGGGTGCTGACGGCGCGCTGCGAGCTACACCAAGAAGAACTTCCCTGGAG 60
Db 82 ATGGGGTGGGTGCTGACGGCGCGCTGCGAGCTACACCAAGAAGAACTTCCCTGGAG 141
QY 61 AGCTCAAGCTCAGCAGCAGTGAAGGATGACCTGACCGTTCTGCGCCATATGTTGCGG 120
Db 142 AGCTCAAGCTCAGCAGCAGTGAAGGATGACCTGACCGTTCTGCGCCATATGTTGCGG 201
QY 121 AGCAAGAAGCGCGATGATACGCTGCTGCGCTGAGAGCTTCTACGGGCCCCAGGCGCT 180
Db 202 AGCAAGAAGCGCGATGATACGCTGCTGCGCTGAGAGCTTCTACGGGCCCCAGGCGCT 261
QY 181 GCCT-----TT 186
Db 262 GCCTATGATGCTTTCCGGTCGGGTTCCTCTGGGGTGGAGGCCCATGCTCGCTGAGTT 321
QY 187 GCTGCCCGCGCTGGCGGAGCGCTCGAACCCTCATCTGGGTTGACCTGGGTGTTGGCACTGGG 246
Db 322 GCTGCCCGCGCTGGCGGAGCGCTCGAACCCTCATCTGGGTTGACCTGGGTGTTGGCACTGGG 381
QY 247 GAGAAATGTCGATATGATGGCTGATTACATCGACTGGCGAAGTTCAAGTCCATCTACGTG 306
```

```
Db 382 GAGAAATGTCGATATGATGGCTGATTACATCGACTGGCGAAGTTCAAGTCCATCTACGTG 441
QY 307 GTCGACCTGTGCCACTCGCTGTGCGG 331
Db 442 GTCGACCTGTGCCACTCGCTGTGCGG 466
```

Search completed: March 14, 2006, 01:11:07
Job time : 7968 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: March 14, 2006, 01:55:41 ; Search time 131 Seconds
(without alignments)
3489.942 Million cell updates/sec

Title: US-10-620-914-45
Perfect score: 3463
Sequence: 1 MGSGRDRPASVTKKNFSL.....RVNYSFYMARKGAKKON 648

Scoring table:
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Uniprot 05.80.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3441	99.4	666	Q6DN05_CHLRE	Q6dn05 chlamydomon
2	953	27.5	835	Q4IPW6_GIBZE	Q4ipw6 gibberella
3	924	26.7	944	Q4PFF5_USTWA	Q4pff5 ustilago ma
4	910	26.3	741	Q6CE63_YARLI	Q6ce63 yarrowia li
5	891	25.7	752	Q59TS3_CANAL	Q59ts3 candida alb
6	890.5	25.7	790	Q6CJ18_KLUFA	Q6cj18 kluyveromyc
7	881.5	25.5	832	Q52GF6_MAGGR	Q52gf6 magnaporthe
8	881	25.4	820	Q55SU0_CRYNE	Q55su0 cryptococcu
9	864.5	25.0	763	Q6BK65_DEBHA	Q6bk65 debaryomyce
10	850.5	24.6	908	Q7SGV5_NEUCR	Q7sgv5 neurospora
11	845.5	24.4	797	Q5KGX9_CRYNE	Q5kgx9 cryptococcu
12	806.5	23.3	831	Q4X175_ASPFU	Q4x175 aspergillus
13	752	21.7	790	Q5B0L6_EMENI	Q5b0l6 aspergillus
14	458	13.2	415	Q7UYV8_RHOBA	Q7uyv8 rhodopirell
15	409.5	11.8	251	Q7UYV9_RHOBA	Q7uyv9 rhodopirell
16	380.5	11.0	407	Q6NBV1_RHOBA	Q6nbv1 rhodopseu
17	379.5	11.0	416	Q93TQ1_RHOSH	Q93tq1 rhodobacter
18	326.5	9.4	416	Q8UDK6_AGRY5	Q8udk6 agrobacteri
19	321	9.3	432	Q98K98_RHILO	Q98k98 rhizobium 1
20	292.5	8.4	416	Q92NK6_RHIME	Q92nk6 rhizobium m
21	161	4.6	367	Q6MIJ9_BDEBA	Q6mij9 bdellovibri
22	159	4.6	377	Q5YZV2_NOCFA	Q5yzv2 nocardia fa
23	123	3.6	2283	DPOE1_MOUSE	Q5yzv7 mus musculu
24	122	3.5	2283	Q80XH7_MOUSE	Q80xh7 mus musculu
25	120	3.5	221	Q92NK5_RHIME	Q92nk5 rhizobium m
26	120	3.5	451	CLUS_COTJA	P14018 coturnix co
27	120	3.5	716	Q8CD66_MOUSE	Q8cd66 mus musculu
28	119	3.4	983	Q59EA9_HUMAN	Q59ea9 homo sapien
29	119	3.4	2243	Q9UNF3_HUMAN	Q9unf3 homo sapien
30	119	3.4	2286	DPOE1_HUMAN	Q97864 homo sapien
31	119	3.4	2286	Q9Y5S4_HUMAN	Q9y5s4 homo sapien

32	119	3.4	2297	2	Q9Y5S5_HUMAN	Q9y5s5 homo sapien
33	114.5	3.3	1128	2	Q5F3Q4_CHICK	Q5f3q4 gallus gall
34	114.5	3.3	1920	2	Q6BNV7_DEBHA	Q6bnv7 debaryomyce
35	112.5	3.2	672	2	Q8VUQ5_PSEPU	Q8vug5 pseudomonas
36	112.5	3.2	672	2	Q88PN6_PSEPU	Q88pn6 pseudomonas
37	111	3.2	477	2	Q6NCK7_RHOBA	Q6nck7 rhodopseu
38	110.5	3.2	782	2	Q4JCH8_SULAC	Q4jch8 sulfolobus
39	109.5	3.2	236	2	Q7NSYS_CHRVO	Q7ngys chromobacte
40	109.5	3.2	853	2	Q92MT9_RHIME	Q92mt9 rhizobium m
41	109	3.1	198	2	Q65GZ7_BACID	Q65gz7 bacillus li
42	108.5	3.1	222	2	Q8TN85_METAC	Q8tn85 methanosarc
43	108.5	3.1	245	2	Q6NBV2_RHOBA	Q6nbv2 rhodopseu
44	108	3.1	255	2	Q6NTG7_HUMAN	Q6ntg7 homo sapien
45	108	3.1	676	2	Q64QR3_BACPR	Q64qr3 bacteroides

ALIGNMENTS

RESULT 1
Q6DN05_CHLRE PRELIMINARY; PRT; 666 AA.
ID Q6DN05; CHLRE
AC Q6DN05;
DT 25-OCT-2004 (T-EMBLrel. 28, Created)
DT 25-OCT-2004 (T-EMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (T-EMBLrel. 28, Last annotation update)
DE Betaine lipid synthase.
GN Name=BTA1;
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;
OC Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=CC125;
RX PubMed=15701786; DOI=10.1128/EC.4.2.242-252.2005;
RA Riekhof W.R., Sears B.B., Benning C.;
RT "Annotation of Genes Involved in Glycerolipid Biosynthesis in
Chlamydomonas reinhardtii: Discovery of the Betaine Lipid Synthase
BTA1Cr.";
RL Eukaryot. Cell 4:242-252(2005).
DR EMBL: AY656806; AAT72342.1; -, mRNA.
DR GO: GO:0008757; F:S-adenosylmethionine-dependent methyltransf. ...; IEA.
DR GO: GO:0016740; F:transferase activity; IEA.
DR InterPro: IPR002114; HPR Serp S.
DR InterPro: IPR000051; SAM_bind.
DR PROSITE: PS00589; PTS HPR SER; UNKNOWN_1.
KW Methyltransferase; Transferase.
SQ SEQUENCE 666 AA; 75780 MW; 4B6AF5718682284 CRC64;

Query Match 99.4%; Score 3441; DB 2; Length 666;
Best Local Similarity 97.1%; Pred. No. 6.6e-273;
Matches 647; Conservative 1; Mismatches 0; Indels 18; Gaps 1;

QY	1	MGSGRDRPASVTKKNFSL	KLSSMKD	DLTVLRHMWFGSKGDDHAARLESFYGPQAA	60
DB	1	MGSGRDRPASVTKKNFSL	KLSSMKD	DLTVLRHMWFGSKGDDHAARLESFYGPQAA	60
QY	61	AF	-----	AARLAERSNLIWDLGGGTGENVDMADYIDLAKPKSIYV	102
DB	61	AYDAFRSFLWGRRPMLAAVAARLAERSNLIWDLGGGTGENVDMADYIDLAKPKSIYV	120		
QY	103	VDLCHSLCEVAKKAKAKGKNVQVVEADACQAPPEGTATLITFSYSLTMIPIFFHNVID	162		
DB	121	VDLCHSLCEVAKKAKAKGKNVQVVEADACQAPPEGTATLITFSYSLTMIPIFFHNVID	180		
QY	163	QACSYLSQDLGVGVADFYVSGKYDPLRQMPMSRRFFWRSIFDIDNIDIGPERRAYLEQK	222		
DB	181	QACSYLSQDLGVGVADFYVSGKYDPLRQMPMSRRFFWRSIFDIDNIDIGPERRAYLEQK	240		
QY	223	LERVWEQNTQGSIPYVWLRAPYYVWIGRLPSVGHALHEERVERPMPFPPTLYTQSWED	282		
DB	241	LERVWEQNTQGSIPYVWLRAPYYVWIGRLPSVGHALHEERVERPMPFPPTLYTQSWED	300		

QY 283 PEDMEVMEINPKDVTLTGSGCNALNLVQAGQVVSVDNCPAQSALLELKKVAIQOL 342
 DB 301 PEDMEVMEINPKDVTLTGSGCNALNLVQAGQVVSVDNCPAQSALLELKKVAIQOL 360
 QY 343 EFDVWQVLFGEVHPRIEELYEKKLAPFLSOTSHNFWMSKELMTFOHGLYYQGGMGKLCWV 402
 DB 361 EFDVWQVLFGEVHPRIEELYEKKLAPFLSOTSHNFWMSKELMTFOHGLYYQGGMGKLCWV 420
 QY 403 LQCLAVVLGLGKTVKRLANAPTMEEQRLDWSNMLHFVXNGPKPLVWLFVKSVLVFN 462
 DB 421 LQCLAVVLGLGKTVKRLANAPTMEEQRLDWSNMLHFVXNGPKPLVWLFVKSVLVFN 480
 QY 463 KAVLWFGGGVPGKQYALIKADGPIENYIARTWDGVAENSHVRKONYFYNCITGKFLRD 522
 DB 481 KAVLWFGGGVPGKQYALIKADGPIENYIARTWDGVAENSHVRKONYFYNCITGKFLRD 540
 QY 523 NCPTYLREAAAFATLKSGVVDNLTVSTNFFMEELKARTYTKVILMDHVDLMDMPVANELAE 582
 DB 541 NCPTYLREAAAFATLKSGVVDNLTVSTNFFMEELKARTYTKVILMDHVDLMDMPVANELAE 600
 QY 583 CLAKQVAPGGIVWRSASLSPVAELIQKAGFVRCIRRTATQGMORVNMYSFYMARK 642
 DB 601 CLAKQVAPGGIVWRSASLSPVAELIQKAGFVRCIRRTATQGMORVNMYSFYMARK 660
 QY 643 GAKKON 648
 DB 661 GAKKON 666

RESULT 2

Q4IPW6_GIBZE PRELIMINARY; PRT; 835 AA.
 ID Q4IPW6_GIBZE PRELIMINARY; PRT; 835 AA.
 AC Q4IPW6;
 DT 13-SEP-2005 (TrEMBLrel. 31, Created)
 DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
 DE Hypothetical protein.
 GN Q4IPW6_GIBZE
 OS Gibberella zeae PH-1.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Eupocreomycetidae; Hypocreales; Nectriaceae; Gibberella.
 OX NCBI_TaxID=229533;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=PH-1;
 RA Birren B., Nusbaum C., Abouelleil A., Allen N., Anderson S.,
 RA Arachchi H.M., Barna N., Bastien V., Bloom T., Boguslavskiy L.,
 RA Boukhalter B., Butler J., Calvo S.E., Camarata J., Chang J.,
 RA Choepe Y., Collymore A., Cook A., Cooke P., Corum B., DeArellano K.,
 RA Diaz J.S., Dodge S., Dooley K., Dorris L., Elkins T., Engels R.,
 RA Erickson J., Faro S., Ferreira P., Fitzgerald M., Gage D., Galagan J.,
 RA Gardyna S., Gnerre S., Graham L., Grand-Pierre N., Hafez N.,
 RA Hagopian D., Hagos B., Hall J., Horton L., Hulme W., Iliev I.,
 RA Jaffe D., Johnson R., Jones C., Kamal M., Kamat A., Karatas A.,
 RA Kells C., Landers T., Levine R., Lindblad-Toh K., Liu G., Lui A.,
 RA Ma L.-J., Mabbitt R., Maclean C., Macdonald P., Major J., Manning J.,
 RA Matthews C., Mauceli E., McCarthy M., Meldrim J., Meneus L.,
 RA Mihova T., Mienga V., Murphy T., Naylor J., Nguyen C., Nicol R.,
 RA Nielsen C.B., Norbu C., O'Connor T., O'Donnell P., O'Neill D.,
 RA Oliver J., Peterson K., Phunkhang P., Pierre N., Purcell S.,
 RA Rachupka A., Ranasingh U., Raymond C., Retta R., Rise C., Rogov P.,
 RA Roman J., Schauer S., Schupbach R., Seaman S., Severy P., Smirnov S.,
 RA Smith C., Spencer S., Stange-Thomann N., Stojanovic N., Stubbs M.,
 RA Talamas J., Testaye S., Theodore J., Topham K., Travers M.,
 RA Vassiliev H., Venkataraman V.S., Viel R., Vo A., Wang S., Wilson B.,
 RA Wu X., Wyman D., Young G., Zainoun J., Zembek L., Zimmer A., Zody M.,
 RA Lander E.
 RT "Fusarium graminearum genome sequence."
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
 CC -!- CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 CC preliminary data.

DR EMBL; AACM01000035; EAA70688.1; -; Genomic_DNA.
 KW Hypothetical protein.
 SQ SEQUENCE 835 AA; 93926 MW; 8192222A79ABC056E CRC64;
 Query Match 27.5%; Score 953; DB 2; Length 835;
 Best Local Similarity 32.7%; Pred. No. 8.7e-69;
 Matches 254; Conservative 103; Mismatches 216; Indels 204; Gaps 25;
 QY 40 GSKGDDHAALESFYGPQAAAF-----AARLA-----ERSNLIV 75
 DB 64 GSKGQQA--LESFYKQAGDAYDTRKVLRLGREDMLALVAAQMAKLDKNKTKRVV 122
 QY 76 DLGGGTGENVDWADYIDLAK-FKSIYVVDLCHSLCEVAKKAKAKGKNVQVVEADACQ 134
 DB 123 DIGGGTGFNIEAMGAFVDVPTFFSSVYLVDSPSLCEVAKKRPDLRGKKNVQVVEADACQ 182
 QY 135 F-----APPE-----GTALITFSYSLTWIPFPHNVIDOACS 167
 DB 183 FRLEDYESGMPKSIIPRSPALSYFDKPRDFGADLITMSYSLSMIPDYISVIDSVTSL 242
 QY 168 LSQDGLGVGADVYVSGKYDLP-----RQMPWRRFFWRSIFDIDNIDIGPERRAY 218
 DB 243 LSPQGGVGVDFVQNKVDFAFRNYTGGVLDVRHVNFLSRFSWRSWFDLDRVGLPERRDY 302
 QY 219 LEQKLERVWEQNTQ-----GSIPYVWLRAPYVWIG--RLPSYGHALHEEVER----- 266
 DB 303 LEYKFGTVLVNTRNKGAL-----PYIWLGHCKKPPSSSLPHEIVERIDALVT 354
 QY 267 --PPMFP-----PTFLY----- 276
 DB 355 ESPLYPANHGDAITRAISAPERSKAFLTAVSNLSSNLPSPFPYQNHWRIVYDEQ 414
 QY 277 ----TQ-----SWDEPEDMEVMEINPKDVTLTGSGCNALNLVQAGQVVSVD 323
 DB 415 LPKHTQFKDEIYIAFTVEDTRVDERILKLGADDKVLAITSGAGNIILSYLQSPARVHVD 474
 QY 324 CNPAQSALLELKKVAIQOLEFEDVWQVLFGEVHPRIEELYEKKLAPFLSOTSHNFWMSKRL 383
 DB 475 LNPTQNHLELKAASYTALPYGDFWKIFGDKHPRFRELLITKLSPHLSGRAFYQWLNK 534
 QY 384 WYFQ-----HGLYYQGGMGKLCWVLOCLAVVLGLGKTVKRLANAPTMEEQRLDWSNMLIH 439
 DB 535 HVPQSSGYGLYDTGGSRAIRVFRWIARIFGLQKAVKQLLQAKTLNEQREIWRRI--- 591
 QY 440 FVKNQPKLVWLFVKFV-SLVLENKAVLWFGGVPGKQYALIKAD----- 483
 DB 592 -----RPA--LUSKLVNCLVVSQESFLWALGVPKNQLAMIEDHANSDLVKGPKPAK 643
 QY 484 ---GIPENYIARTMDGVAENSHVRKONYFYNCITGKFLRDNCPTYLREAAAFATL-KSG 539
 DB 644 NTRSHALWHYVWNTLPVAAETHIAADNPYYVYVMDGKFPKCHPDYLSPRAHAKLSRP 703
 QY 540 VDNLTVSTNFFMEELKAR-----TYTKVILMDHVDLMDMPVANELAE--LAKQVAPGGI 593
 DB 704 ALDGLAIHTD-ELEEVYIARTITPGTLTIVAVVMDMSMDPDTGSRAAAQITKLNALAMGGR 762
 QY 594 VVRSASLSPVAELIQKAGF-----DVRCIRRTATQGMORVNMYSFYMARK 641
 DB 763 VLLRSSALTWPYIKBPEGHGFTPKRHGARIDGACI-----DRVNMYASCIWCTK 811
 RESULT 3
 Q4IPW6_USTMA
 ID Q4IPW6_USTMA PRELIMINARY; PRT; 944 AA.
 AC Q4IPW6;
 DT 13-SEP-2005 (TrEMBLrel. 31, Created)
 DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
 DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
 DE Hypothetical protein.
 GN ORFNames=UM01159.1;
 OS Ustilago maydis 521.
 OC Eukaryota; Fungi; Basidiomycetes; Ustilaginomycetes;
 OC Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.

OX NCBI_TaxID=237631;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=521;
 RA Birren B., Nusbaum C., Abebe A., Abouelleil A., Adekoya E.,
 RA Alit-zahra M., Allen N., Allen T., An P., Anderson M., Anderson S.,
 RA Arachi H., Ambruster J., Bachanteang P., Baldwin J., Barry A.,
 RA Bayul T., Blithsteyn B., Bloom T., Blye J., Boguslavskiy L.,
 RA Borowsky M., Boukhalter B., Brunache A., Butler J., Calixte N.,
 RA Calvo S., Camarata J., Campo K., Chang J., Cheshatsang Y., Citroen M.,
 RA Collimore A., Considine T., Cook A., Cooke P., Corum B., Cuomo C.,
 RA David R., Dawoe T., Degray S., Dodge S., Dooley K., Dorje P.,
 RA Dorjee K., Dorris L., Duffey N., Dupes A., Ekins T., Engels R.,
 RA Erickson J., Farina A., Faro S., Ferreira P., Fischer H.,
 RA Fitzgerald M., Foley K., Gage D., Galagan J., Gearin G., Gnerre S.,
 RA Gnirke A., Goyette A., Graham J., Grandbois E., Gyaltsen K., Hafez N.,
 RA Hagopian D., Hagos B., Hall J., Hatcher B., Heller A., Higgins H.,
 RA Honan T., Horn A., Houde N., Hughes L., Hulme W., Husby E., Iliev I.,
 RA Jaffe D., Jones C., Kamal M., Kamat A., Kamysseleis M., Karlsson E.,
 RA Kells C., Kieu A., Kienner P., Kodira C., Kulbokas E., Labutti K.,
 RA Lama D., Landers T., Leger J., Levine S., Lewis D., Lewis T.,
 RA Lindblad-toh K., Liu X., Lokitsang T., Lokitsang Y., Lucien O.,
 RA Lui A., Ma L.J., Mabbitt R., Macdonald J., Maclean C., Major J.,
 RA Manning J., Marabella R., Maru K., Matthews C., Mauceli E.,
 RA McCarthy M., McDonough S., Mcghee T., Meldrim J., Meneus L.,
 RA Mesirov J., Mihaliev A., Mihova T., Mikelsen T., Mienga V., Moru K.,
 RA Mozes J., Mulraiv L., Munson G., Naylor J., Newes C., Nguyen C.,
 RA Nguyen N., Nguyen T., Nicol R., Nielsen C., Nizzari M., Norbu C.,
 RA Norbu N., O'donnell P., Okoawo O., O'leary S., Omotosho B.,
 RA O'Neill K., Osman S., Parker S., Perrin D., Phunkhang P., Pignani B.,
 RA Purcell S., Rachupka T., Ramasany U., Rameau R., Ray V., Raymond C.,
 RA Retta R., Richardson S., Rise C., Rodriguez J., Rogers J., Rogov P.,
 RA Rutman M., Schubach R., Seaman C., Settipalli S., Sharpe T.,
 RA Sheridan J., Sherpa N., Shi J., Smirnov S., Smith C., Sougnez C.,
 RA Spencer B., Stalker J., Stange-chomann N., Stavropoulos S.,
 RA Stetson K., Stone C., Stone S., Stubbs N., Talamas J., Tchuanga P.,
 RA Tenzon P., Tesfaye S., Theodore J., Thoulutsang Y., Topham K.,
 RA Towey S., Tsamila T., Tsomo N., Vallee D., Vassiliev H.,
 RA Venkataraman V., Vinson J., Vo A., Wade C., Wang S., Wangchuk T.,
 RA Wangdi T., Whittaker C., Wilkinson J., Wu Y., Wyman D., Yadav S.,
 RA Yang S., Yang X., Yeager S., Yee E., Young G., Zainoun J., Zembeck L.,
 RA Zimmer A., Zody M., Lander E.;
 RT "The genome sequence of *Ustilago maydis*.";
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
 CC -!- CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 DR EMBL; AACP0100043; EAK81682.1; -; Genomic_DNA.
 DR InterPro; IPR000051; SAM_bd.
 KW Hypothetical protein; Methyltransferase; Transferase.
 SQ SEQUENCE 944 AA; 106525 MW; B7C3A32D89132B0D CRC64;

 Query Match 26.7%; Score 924; DB 2; Length 944;
 Best Local Similarity 30.0%; Pred. No. 2.4e-66;
 Matches 234; Conservative 113; Mismatches 237; Indels 196; Gaps 22;

 QY 43 KGDHARLESFYGPQAAAF-----AARL--ABRSN-----LIWVDLG 78
 DB 179 KTAQAEALDKFFQONQANVYDSTRGLKGRRTTMLKCAQLRDMQASNPGKPLWVDIG 238

 QY 79 GGTGENDVMADYIDIAKPKSIYVVDLCHSLCEVAKKAKAKGKWKVQVVEADACQFAPP 138
 DB 239 GGTGWNIEQWQFPFDQLSQVYLIDLCPLQAVARKRAAKGKVKVQLCQDASQFNMP 298

 QY 139 EGTAT-----LITFSLSLTIWPPHNVIDQACSYLSQ-DGLGVADFYVSGKVDLPL--- 189
 DB 299 -GLAAGQKVDLFTCSYSIMPPFYAVLDRINDLDPVTGVGVDFYVSGSG-PLAKS 356

 QY 190 -----ROMPSRRFFWRSIFDIDNIDIGPERRAYLEOKLVRVWEQNTQGS--IPYVP 239
 DB 357 PMTGGTRQCQGLSWRFWSMFWSPFDHIELHPARDYLEHKEFTIKCYNGRNNFIPTF- 415
 QY 240 WLRAPYVWIG--RLPSVGHALHEERVER-----PPMPPP----- 272

Db 416 -VRPIYIWLGVSRERDRTTKAIQAIFESGNGRVVVPSPPELAYMNGEAHTHTSIAAT 474
 QY 273 ----- 272
 Db 475 ATATATATGIDADEDASSRPEPARRLLRRVSEATTESDADSDRPLKLELGHPPPLSSPH 534
 QY 273 -----TFLYTQSWEDPEPDMVMEINPKDVTVLTLTSGGCNAINLL 312
 Db 535 YQKRWLPFVDFNEFSDMERTWYGTWEDPYVDMQHLGLKDDSLCITSAGDNALHYA 594
 QY 313 VQG-AGQVVSVDNCPAQSALELKKVAIQOLEPQEDVQVLFEGVHPRIEELYEKLAAPFL 371
 Db 595 VAKPRRIHADVMPCCGHLLELKLACIASLSDYDMMQMPGEGRIDNFRLELLDSKISPYL 654
 QY 372 SQTSHNFWSKRLWYFQGLYVQCGMKLQWLQCLAVLGLGKTVKRLANAPTMEORRL 431
 Db 655 SSHAVQFWRLNTRAFDKAFYPRGYSCHALRLAKAFSVTVGRVWEKMTCTANSVEQEV 714
 QY 432 WDSNMLIHVFKNGPKPLVWLFKVFLVFNKAVLWFGGVPKGVKQYALIKADGPIENYI 491
 Db 715 WDKCLRSTLIN---KPLIRLF-----LSNPAFLWNLGVPNMNQIIFLNEGVSABQFA 764
 QY 492 ARTMDGVAENSHVRKQNYFYNNCLTKGFLRDNCPYTLREAAAFATIKSGV-----VDNLTVS 547
 Db 765 IDTLDSIPRSRLTKNDNYHQYCLLHKYTKQSCPLYLKPDPGFAALKQALQDGLDSFRLH 824
 QY 548 TNPFMBELKA---RTYTKVILMDHVDWLD-MPVA-----NELAE 582
 Db 825 TDSIVNVLGFDPGALTRATWDMDFDPVPASRPAPTIKQARRDSKSVSDLDREICE 884
 QY 583 CIAQVAPGGIVWIRASISLSPYAEILIKAGFDVRCIR-RATQGYMDRVNMVSSFYMAR 641
 Db 885 -LSRVIRKGVAFYRSAAKPKWYHRRFKMGFSQVPHIRETAKPIDNVNMYASFYKATR 943

 RESULT 4
 ID Q6C6E3 YARLI PRELIMINARY; PRT; 741 AA.
 AC Q6C6E3;
 DT 25-OCT-2004 (TrEMBLrel. 28, Created)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
 DE Similar to DEHA0F25982g Debaryomyces hansenii IPF 6093.1.
 GN OrderedLocustNames=YALIOE10197g;
 OS Yarrowia lipolytica (Candida lipolytica).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Dipodascaceae; Yarrowia.
 OX NCBI_TaxID=4952;
 RN [1]
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
 RC STRAIN=CLIB 122 / E 150;
 RX PubMed=15229592; DOI=10.1038/nature02579;
 RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
 RA Lafontaine I., de Montigny J., Marck C., Neuvéglise C., Talla E.,
 RA Goffard N., Frangeul L., Aigle M., Anchoard V., Babour A., Barbe V.,
 RA Barnay S., Blanchin S., Beckerich J.-M., Beyne E., Bleykasten C.,
 RA Boisrame A., Boyer J., Cattolico L., Confanioleri F., de Daruvar A.,
 RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
 RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
 RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
 RA Nicaud J.-M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
 RA Pellenz S., Potier S., Richard G.-P., Straub M.-L., Suleau A.,
 RA Swennen D., Tekia F., Wesolowski-Louvel M., Westhof E., Wirth B.,
 RA Zenlou-Meyer M., Zivanovic Y., Bolotin-Fukuhara M., Thierry A.,
 RA Bouchier C., Caudron B., Scarpelli C., Gallardin C., Weissenbach J.,
 RA Wincker P., Souciet J.-L.;
 RT "Genome evolution in yeasts.";
 RL Nature 430:35-44 (2004).
 DR EMBL; CR382131; CAG79360.1; -; Genomic DNA.
 DR GO; GO:0008757; F.S-adenosylmethionine-dependent methyltransferase; IEA.
 DR GO; GO:0016740; F:transferase activity; IEA.
 DR InterPro; IPR000051; SAM_bind.


```

Db 704 LRSASTKFWYLTGKTLGFOEBEENVRVQPGSSIDRVNMYAN 744
Q6CJ18_KJULA
ID Q6CJ18_KJULA PRELIMINARY; PRT; 790 AA.
AC Q6CJ18; 2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Similar to ca|C1783|pF7635 Candida albicans unknown function.
GN OrderedLocName=KJLA0F22198;
OS Kluyveromyces lactis (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
OX NCBI_TaxID=28985;
RN [1];
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=CBS 2359 / IFO 1267 / NRRL Y-1140 / WM37;
RX PubMed=15229592; DOI=10.1038/nature02579;
RA Lafontaine I., de Montigny J., Marck C., Neugebäude C., Talla E.,
RA Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.-M., Beyne E., Bleykasten C.,
RA Boisrame A., Boyer J., Cartolico L., Confanioli F., de Daruvar A.,
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Nicaud J.-M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
RA Pellenz S., Potier S., Richard G.-P., Straub M.-L., Suteau A.,
RA Swennen D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zentou-Meyer M., Zivanovic Y., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA Wincker P., Souciet J.-L.;
RT "Genome evolution in yeasts.";
RL Nature 430:35-44(2004).
DR EMBL; CR382126; CAG98779.1; -; Genomic DNA.
DR GO; GO:0008757; F:adenosylmethionine-dependent methyltransferase; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000051; SAM_bind.
KW Complete proteome; Methyltransferase; Transferase.
SQ SEQUENCE 790 AA; 90212 MW; 25BAE985E9C5B0D1 CRC64;

Query Match 25.7%; Score 890.5; DB 2; Length 790;
Best Local Similarity 33.2%; Pred. No. 1.1e-63;
Matches 233; Conservative 94; Mismatches 241; Indels 133; Gaps 23;

QY 51 LRSFYCPQAAAP-----AARLAERSNLIWDLGGTGGENVDMADYI 92
DB 93 LEQFYKSOAKLYDRTRGVLLQGRSTSLKLSHLSEKGNWIDVGGTGFNISQWALIT 152
QY 93 DL-AKFSIYVVDLCHSLCEVAKKAKGKNNQVVEADACAPAPPEGTATLITFSYSL 151
DB 153 NLDTPDKYLLIDLSPLCEVAKKCKEKGKNNVEICGDACDFELPEESAQILITFSYSL 212
QY 152 TMIPPHNVIDQACSYL-SQDGLVGVADFYVSGKY-----DLPLRQMPWSRRFFW 200
DB 213 SMIPSPYALDHAVSLLDKAGIISCVDFGVNESMLVGRWTGLGVNHRHLPWLPRTW 272
QY 201 RIFIDNIDIGERRAYLEOKLERVWEQNTQ-----GSIPYVPWL-----241
DB 273 RLWFEFDKFLDPARREYLEYRFGTIKSLNCYNYKLGKIPYITWLGCKNDHEQHLQARFV 332
QY 242 ----RAPPYVWTCR-----LPSVGHALHEERV-----ERPP- 268
DB 333 ELAPTSYPLAPITTSASSNAQPMWKAMIAALENSKKGLPYSLFYQKEHWRVYDEVNPE 392
QY 269 --MFPPTFLYTOSWEPDEPMVEINPKDTVLITLTSGCCNALN--LLVQAGAGVVSVD 324
DB 393 YSQFNKSYIYATWEDPREDVNIILNQTEDTILATISAGNLIHLYATLPNPKRTHGVDL 452
QY 325 NPAQSALLBKVAIQOLEFFEDVQWLFGEVHPRIEELYEKKLAPPLSQTSHNFWSKRLW 384

```

```

Db 453 NPCQGHLELKLAAIRSLSTFTQWQMGEGKIDRFNNILNKLAPYLSSNAFYW----- 507
QY 385 YFOHGL-----YYQGMGKLCWVLOC---LAVVLGLGKTVKRLANAPTMEORRLW 432
DB 508 -FENGTKTFDPNGAGLYDTGFTK--WALRLAKWVFKVANLTDEVNMLCKAKTIEEQRSIW 564
QY 433 DSNMLIHFKVNGPKPLVWLFVKFVSLVLF-NKAVLMFGGVGPKGVKVALIKADGPIENYI 491
DB 565 DKKI-----KPV--LFRNVGKILVGNPLFLMSALGVPRNQ---AKMGSSSTLOYI 610
QY 492 ARTMDGVAENSHVRKONYFYVYNTCLTKGFLRDNCPYTLREAAFPATLK-----SGVVDNLTVS 547
DB 611 IDTLDVIDNLSLDDNYFYTLTKGRYSRSCPDYKKGFGKSLSRSPESPLDVRHL 670
QY 548 TNFF---MEELKARTYTKVILMDHVDWLDMPVANELAE---CLAKQVAPGGIWIWRSASL 601
DB 671 TDTLKDVCERLSKKTVSIALIMHDMDFD-POGTDVDEIQTALMLNLSRGRVLLRSASK 729
QY 602 SPYAEIOLKAGFDVRCIRRTAQG-YMDRVNMYSSFYMARR 641
DB 730 SPWYIKNFKFGFSCKAVSARYPGKICIDRVNMYASTWVCQK 770

RESULT 7
Q52GF6_MAGGR
ID Q52GF6_MAGGR PRELIMINARY; PRT; 832 AA.
AC Q52GF6;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DE Hypothetical protein.
GN ORFNames=MG01330.4;
OS Magnaporthe grisea 70-15.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetes incertae sedis; Magnaportheaceae; Magnaporthe.
OX NCBI_TaxID=242507;
RN [1];
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=70-15;
RA Birren B., Nusbaum C., Abebe A., Abouelleil A., Adekoya E.,
RA Ait-zahra M., Allen C., Allen T., An P., Anderson M., Anderson S.,
RA Arachchi H., Armbruster J., Bachantsang P., Baldwin J., Barry A.,
RA Bayul T., Blitshsteyn B., Bloom T., Blye J., Boguslavskiy L.,
RA Borowsky M., Boukhgalter B., Brunache A., Butler J., Calixte N.,
RA Calvo S., Camarata J., Campo K., Chang J., Cheshatsang Y., Cicroen M.,
RA Collymore A., Considine T., Cook A., Cooke P., Corum B., Cuomo C.,
RA David R., Dawoe T., Degray S., Dodge S., Dooley K., Dorje P.,
RA Dorjee K., Dorris L., Duffey N., Dupes A., Elkins T., Engels R.,
RA Erickson J., Farina A., Faro S., Ferreira P., Fischer H.,
RA Fitzgerald M., Foley K., Gage D., Galagan J., Gearin G., Gnerre S.,
RA Gnitke A., Goyette A., Graham J., Grandbois E., Gyaltsen K., Hafez N.,
RA Hagopian D., Hagos B., Hall J., Hatcher B., Heller A., Higgins H.,
RA Honan T., Horn A., Houde N., Hughes L., Hulme W., Husby E., Iliev I.,
RA Jaffe D., Jones C., Kamal M., Kamat A., Kamysseis M., Karlsson E.,
RA Kella C., Kieu A., Klsner P., Kodira C., Kulbokas E., Labutti K.,
RA Lama D., Landers T., Leger J., Levine S., Lewis D., Lewis T.,
RA Lindblad-toh K., Liu X., Lokyitsang T., Lokyitsang Y., Lucien O.,
RA Lui A., Ma L.J., Mabbitt R., Macdonald J., Maclean C., Major J.,
RA Manning J., Marabella R., Maru K., Matthews C., Mauceli E.,
RA McCarthy M., McDonough S., Mcghee T., Meldrum J., Meneus L.,
RA Mesirov J., Mihalev A., Mihova T., Mikelsen T., Mienga V., Moru K.,
RA Mozes J., Mulrain L., Munson G., Naylor J., Neues C., Nguyen C.,
RA Nguyen N., Nguyen T., Nicol R., Nielsen C., Nizzari M., Norbu C.,
RA Norbu N., O'donnell P., Okoawo O., O'leary S., Omotosho B.,
RA O'Neill K., Oman S., Parker S., Perrin D., Phunkhang P., Pignani B.,
RA Purcell S., Rachupka T., Ramasamy U., Rameau R., Ray V., Raymond C.,
RA Retta R., Richardson S., Rise C., Rodriguez J., Rogers J., Rogov P.,
RA Rutman M., Schupbach R., Seaman C., Settipalli S., Sharpe T.,
RA Sheridan J., Sherpa N., Shi J., Smirnov S., Smith C., Sougneux C.,
RA Spencer B., Stalker J., Stange-thomann N., Stavropoulos S.,
RA Steenson K., Stone C., Stone S., Stubbs M., Talamas J., Tchuinga P.,
RA Tensing P., Tesfaye S., Theodore J., Thoulutseang Y., Topham K.,
RA Towey S., Tsamla T., Tsomo N., Vallee D., Vassiliev H.,

```

RA Venkataraman V., Vinson J., Vo A., Wade C., Wang S., Wangchuk T.,
RA Wangdi T., Whittaker C., Wilkinson J., Wu Y., Wyman D., Yadav S.,
RA Yang S., Yeager S., Yee E., Young G., Zainoun J., Zembeck L.,
RA Zimmer A., Zody M., Lander E.,
RT "The genome sequence of Magnaporthe grisea."
RT Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=70-15;
RA Dean R., Mitchell T., Brown D., Pan H., Thon M.;
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=70-15;
RA Zhu H., Blackmon B.;
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AACU01000030; EAA55679.1; -; Genomic_DNA.
KW Hypothetical protein; Methyltransferase; Transferase.
SQ SEQUENCE 832 AA; 93328 MW; 98ASCF0B54A618AE CRC64;

Query Match 25.5%; Score 881.5; DB 2; Length 832;
Best Local Similarity 31.3%; Pred. No. 6.3e-63;
Matches 226; Conservative 114; Mismatches 226; Indels 155; Gaps 21;

QY 51 LESFYGPQAAAF-----AARLAERS-----NLIVDLGGGTGENDV 86
DB 67 LESFYETQASIVDVTTLTKGREDMLAALAAQLHRAKQSHQKRIWVDIGGGTGNIE 126
QY 87 MMADYIDLAK-FKSIYVVDLCHSLCEVAKKAKGKNVQVVEADACOF-----APP 138
DB 127 AMRFVDVPEFFSVLVDPSPSLCEVARFORLGNKVRVVCQDARFRIEDHEASGP 186
QY 139 EG-----TATLIFFSYSLTMIPIFFHNVIDQACSYLSQDGLGVADFYVSGKYD----- 186
DB 187 KGNWQAIGADLLTWSYLSLMIPIFYSDLSLLAPGGIMGVVDVFAQSKVDPSFRNHT 246
QY 187 --LPLRQMPWSRRFFWRSIFDINIDIGPERRAYLFQKLERWEQNTQ-----GSIYYPVW 240
DB 247 AGLTGRHCGFILRSFRAWFDLRSLEPARDYLFYRFTGLVGNVTRNRLGNI----- 301
QY 241 LRAPYVWIG-----RLEPSV-----GHA---L 259
DB 302 ---PYIIVICSTKSGQSTQSLHENVSPKAVQVAVKNSAGLPLPSFYQNHAWRIY 358
QY 260 HEBRVERPPMPFTFLYTQSWEDPEPDMEYMEINPKDVTLTLSGGCNALNLLVQAGQV 319
DB 359 YDQLRKHKQFNDEYIYAFWTEDTRVDARLLKISSDDVLAITSAGDNILSYISQSPARV 418
QY 320 VSDCNPAQASALLEKVAIQOLEFEDVQWLFQEGVHPRIEELYEKKLAPFLSQTSHNFW 379
DB 419 HAVDLNPSQNHLLLEKLASYSDIADYDFWKIFGEGHDEFRALLINKLSPLHSRAFYW 478
QY 380 --SKRLWYFGLYQYGGMKLCWVLQCLAVLGLGTVKRLANAPTEEQRLWDSNML 437
DB 479 FANAHTSFKHGYDYGGSKHATRAFRWISSAFGVRSAVREMLEAKTNEQREIWOGR- 537
QY 438 IHFVGNPKPLVMLFVKFVS-LVLFNKAVLWFGGVPKGQYALIKADGI----- 485
DB 538 -----RA--LLSRMVSNFVVSQQLWSALGVPKHQLAMLERDHATSKVVSERGEAS 587
QY 486 -----PIENYIARTMDGVAENSHVRKQNYFYNNCLTKGFLRDNCPYTLREAAAFATL--KS 538
DB 588 NSRTSAIQWMSDITLDPVAHDTTHAENPYVYVYTMAGKFSRCHPDYLSPAAHAKLSRRG 647
QY 539 GVVDNLTVST--NFFMEELKARTYTKVILMDHVDMLDMP---VANELASCLAKQVAPGG 592
DB 648 ALERITRIHTEEDIVVLGAPSLTVAVWSDMSDFDPKPEGVAAAQISKLNRALPKGG 707
QY 593 IIVWSASLSPPAELIQAGF-----DVRCIRRATQGYMDRVNMYSSFYMARRK 642

DB 708 RYLLRSAGLKPWYVEEFAKLGFPAPKRVGCRSPETRCI-----DRVNMVASCWLLTKT 759
QY 643 G 643
DB 760 G 760

RESULT 8
Q55SJ0_CRYNE PRELIMINARY; PRT; 820 AA.
AC Q55SJ0;
DT 13-SEP-2005 (TREMBLrel. 31, Created)
DT 13-SEP-2005 (TREMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TREMBLrel. 31, Last annotation update)
DE Hypothetical protein.
GN ORFNames=CNBEI930;
OS Cryptococcus neoformans var. neoformans B-3501A.
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Heterobasidiomycetes;
OC Tremellomycetidae; Tremellales; Tremellaceae; Filobasidiella.
OX NCBI_TaxID=283643;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=B-3501A;
RA Fung E., Hyman R.W., Rowley D., Bruno D., Miranda M., Fukushima M.,
RA Wickes B.L., Fu J., Davis R.W.;
RT "Cryptococcus neoformans serotype D sequencing."
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAEY01000024; BAL20831.1; -; Genomic_DNA.
KW Hypothetical protein.
SQ SEQUENCE 820 AA; 93328 MW; 0F694988B89353A2 CRC64;

Query Match 25.4%; Score 881; DB 2; Length 820;
Best Local Similarity 29.2%; Pred. No. 6.7e-63;
Matches 223; Conservative 110; Mismatches 232; Indels 198; Gaps 21;

QY 40 GSKKGDHAAARLESFYGPQAAAF-----AARL-----AERS 70
DB 88 GSRQKDH---LEAFVAGQADLYDITRSQLLKGRFTMLQALLAHLKAAQMPVRLVPSAPSK 144
QY 71 NLIVDLGGGTGENDVMADYIDLAKFKSIYVVDLCHSLCEVAKKAKGKNVQVVEA 130
DB 145 PRIWDLGGGTGWNIEKMDYLPITYFDIYIIDLCEPLLEVARARIKARGKNVHLCQ 204
QY 131 DACQAPPGETATLIFFSYSLTMIPIFFHNVIDQACSYL-SODGLGVADFYVVS---GKYD 186
DB 205 DASRFVLPE-----WESGAIPPFYQVLDRCQVLDTPQGLMAVVDFTTGREVNKE 255
QY 187 LPL-----RQMPWSRRFFWRSIFDINIDIGPERRAYLFQKLERWEQNTQGSIPYVPWL 242
DB 256 RAIGTASRVSWSFKWFEWCFDLGVLHLSRREYLEYKMGTIKYNAENNFINTWFIQ 315
QY 243 APYVWIG----- 250
DB 316 IPYVFLGCSQRDASASAKSFTEAGNRLQSDGLGLTPTSPFTNSPSVFGSPSPMLEM 375
QY 251 -----RLEPSVCHALHEERVERPPPTFLY 276
DB 376 PELVLGFSAAQATQTFEVGAPLSPPHYHLRKAWRIPYLEEKEHQF-----RTHIY 427
QY 277 TQSWEDPEPDMEYMEINPKDVTLTLSGGCNALN-LIVQAGAGVSVVDCNPAQSALLELK 335
DB 428 GWTWEDPAVDVKLNKNDHILAIYSAGNVHLVYALTAKPARIHAVDMNPPCQHILLELK 487
QY 336 KVAIQOLEFEDVQWLFQEGVHPRIEELYEKKLAPFLSQTSHNFWSKLWTFQHLGYQGG 395
DB 488 LAALQALEYNDFWLIFGEGRHPEFDKLLTTKLSPFLSSHAYAWYKTKTSQFSRNFYFRGY 547
QY 396 MGKLCWVLQCLAV---VLGLGTVKRLANAPTEEQRLWDSNMLIHFVGNPKPLVWLF 452

Db 548 SG---WALRLAQTAFFIAGVRGDKVRLCQATSTABQRIWQKKI-----RP-VFLN 594

Qy 453 VKFVSLVFNKAVLWFGGGVPGKQYALIKADGIPNIYARTMDGVAENSHVRKQNYFY 512

Db 595 KVMVGLFLGNPLFNWALGVQNMNCFLDQG-SVEDYVKATLDPITLTLTKLDNNYFPL 653

Qy 513 NCLTGKFLRNCCTYTLREAAAFATLK-SGVVDNLTVTSTNFMWELKA---RTYTKVILMDH 568

Db 654 LCLNGRYTTSCEPAFLKPEGFALKRNSKSTDAFKLHTDTTLNLRGLPDBESLTKIIVMS 713

Qy 569 VDWLD-----MPVAN-----ELAECLAKOVAAPGVIVIR 597

Db 714 MDMFIDPCTPDPQDSTALDTLOATPEKALEHLRAELDYELD-MKRVKLVKGGAIVR 772

Qy 598 SASLSPPYAEALIKAGPDVRCIR-RATQGYMDRVNMYSSFYMA 639

Db 773 SAAKRPWYRQRPFAAGLVKQPIDIRENQEAIDRVNMYVSPWKA 815

RESULT 9

ID Q6BK65 DEBHA PRELIMINARY; PRT; 763 AA.

AC Q6BK65;

DT 25-OCT-2004 (T-EMBLrel. 28, Created)

DT 25-OCT-2004 (T-EMBLrel. 28, Last sequence update)

DE Similar to CA1783|IPF7635 Candida albicans IPF7635.

GN OrderedLocusNames=DEHA0F25982g;

OS Debaryomyces hansenii (Yeast) (Torulaspora hansenii).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Debaryomycetes.

OX NCBI_TaxID=4959;

RN [1]

RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].

RC STRATN=ATCC 36239 / CBS 767;

RX PubMed=15229592; DOI=10.1038/nature02579;

RA Dujon B., Sherman D., Fischer G., Durrans P., Casaregola S., Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talla E., Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V., Barnay S., Blanchin S., Beckerich J.-M., Beyne E., Bleykasten C., Boisrame A., Boyer J., Cattolico L., Confanioleri F., de Daruvar A., Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A., Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R., Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H., Nicaud J.-M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O., Pellenz S., Potier S., Richard G.-F., Straub M.-L., Suleau A., Swennen D., Tekia F., Wesolowski-Louvel M., Westhof E., Wirth B., Zeniou-Meyer M., Zivanovic Y., Bolotin-Fukuhara M., Thierry A., Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J., Wincker P., Souciet J.-L.;

RT "Genome evolution in yeasts."

RL Nature 430:35-44 (2004).

DR EMBL: CR382138; CAG89817.1; -; Genomic DNA.

DR GO: GO:0008757; F1S-adenosylmethionine-dependent methyltransf. .; IEA.

DR GO: GO:0016740; F:transferase activity; IEA.

DR InterPro: IPR000223; Peptidase_S26A.

DR InterPro: IPR000051; SAM_bind.

DR PROSITE: PS00501; SPASE_I_1; UNKNOWN 1.

KW Complete proteome; Methyltransferase; Transferase.

SQ SEQUENCE 763 AA; 87363 MW; 7979A2C300525F56 CRC64;

Query Match 25.0%; Score 864.5; DB 2; Length 763;

Best Local Similarity 31.4%; Pred. No. 1.4e-61;

Matches 223; Conservative 105; Mismatches 252; Indels 131; Gaps 20;

Qy 41 SKGDDHARLESFYGPQAAAF-----AARLARSNLINWDLGGGTG 82

Db 57 TTKPTQQQSLELFYKQAHVYDKTRVLLKGRKECLRLATAHLKSKKOLVWVDIGGGTG 116

Qy 83 ENYDMWADYIDLAK-FKSIYVVDLCHSLCEVAKKAKAGKWNQVVEADACQAFPEGT 141

Db 117 SNIEYMDQWMSISKNFAYVLVDLSPLCEVAKKRFADKRWNSVHLVADACDFGIGYBK 176

Qy 142 ATLITSYSLSLTMIPPHNVIDQSCSYLSQDLGVGVADFVYS-----GKYD-----LPLRQ 191

Db 177 ADLVTSYSLSLTMIPPHNVIDQSCSYLSQDLGVGVADFVYS-----GKYD-----LPLRQ 236

Qy 192 MPMSRFFFRSIFDIDNIDIGPERRAYLEQKLERVWEQNTQSGIPVYVWMLRAPPYVWIG- 250

Db 237 IPVWLNFRWIFWADKVFIDLSRRHYLYKFGTINKSLCYN-----KLGRIPYIYVWGC 292

Qy 251 -----RLP-SVGH----- 257

Db 293 DKSSKSSLLNRINCLATESPYLAPSDDKNDIDIPISKGHEAALMNFKNLPYPSIYQY 352

Qy 258 -----ALHERVERPMPPTFLYTQSWEDPDPMEVWEINPKDVTLTLSGGGNALN--L 311

Db 353 EIWRVYFDEIRQYLOFKNQYVYAFWEDPREPHNLKFTSEDTVLAITSAGNILLSYAT 412

Qy 312 LVQAGQVVSVCNPAQSALLEKVAIOOLEPEDVWOLFGEVGHPRIBELYEKKLAPFL 371

Db 413 LPNPPKRIHAVDLNPNQNHLLLEKLASFALSKNQIWSIFGEGKIKNFKDLIVDKLSPHV 472

Qy 372 SQTSHNFW----SKRLWYFQHGYYQGGMGKLCWLQC----LAVVLGLGKTVKRLANAPT 424

Db 473 SSNTFOYMDRGDKTFNPKGKLY---DTGSTRWALRLAKWPKICGITKYVDMLCBCDT 529

Qy 425 MEEQRLLDSNMLIHFKVNGPKPLWLVFKFV-SLVLFNKAVLWFGGVPGKQYALIKAD 483

Db 530 LEEQQRINWNI-----RPT--LFPNPFVSSLLIGNPIFLWKALGVPAQOANMM--- 575

Qy 484 GIPIENYIARTMDGVAENSHVRKQNYFYNCLTGKFLRNCCTYTLREAAAFATL- 536

Db 576 GNSILKYVVDTFDPVTKRSLISDNIFYTLTKGKTPONCPDYLTEKGYKSTTTNRKS 635

Qy 537 KSGVVDNLTVST---NFFMEELKARTYTKVILMDHVDLMDPVPVANELAECLAKQ--VAPG 591

Db 636 KEAPIDNIRLHTDMLNDVFAFLSKSLSTAILMDHMDWDFDNGEDALNEITALKSLNTN 695

Qy 592 GIVVWRASLSPPYAEALIKAGPDVRCIR-RATQGYMDRVNMYSSFYMA 641

Db 696 GRVLLRSASTNPWYIKTFEDLQFTCKAAGIRDTGISIDRINMYASTVWCTK 746

RESULT 10

ID Q7SGY5 NEUCR PRELIMINARY; PRT; 908 AA.

AC Q7SGY5;

DT 01-MAR-2004 (T-EMBLrel. 26, Created)

DT 01-MAR-2004 (T-EMBLrel. 26, Last sequence update)

DT 01-FEB-2005 (T-EMBLrel. 29, Last annotation update)

DE Hypothetical protein (Related to S-adenosylmethionine:diacylglycerol 3-amino-3-carboxypropyl transferase braa).

GN Name=NCU03032.1; Synonyms=B18P24.050;

OS Neurospora crassa.

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.

OX NCBI_TaxID=5141;

RN [1]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=OR74A;

RA Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D., Jaffe D., FitzHugh W., Ma L.-J., Smirnov S., Purcell S., Rehman B., Ekins T., Engels R., Wang S., Nielsen C.B., Butler J., Endrizzi M., Qui D., Ianakiev P., Pedersen D., Nelson M., Washburne M., Seltremnikoff C.P., Kinsey J.A., Braun E.L., Zelter A., Schulte U., DeSouza C.C., Glass L., Orbach M.J., Berglund J., Voelker R., Roy A., Foley K., Naylor J., Thomann N., Barrett R., Gnerre S., Kamal M., Kamysseles M., Mauceli E., Bielke C., Rudd S., Frishman D., Krystofova S., Ramussen C., Metzberg R.L., Perkins D.D., Kroken S., Cogoni C., Macino G., Catchside D., Li W., Pratt R.J., Omani S.A., DeSouza C.C., Glass L., Orbach M.J., Berglund J., Voelker R., Varden O., Plamann M., Seiler S., Dunlap J., Radford A., Aramayo R., Natvig D.O., Alex L.A., Mannhaupt G., Ebbole D.J., Freitag M., Paulsen I., Sachs M.S., Lander E.S., Nusbaum C., Birren B.;

RT "The Genome Sequence of the Filamentous Fungus Neurospora crassa."

QY 277 TQSWEDPEPMEVMEINPKDTVLTSLTSGCCNALN-LLVQAGQGVSVDCNPAQSALLELK 335
 Db 428 GWTWEDPAVDVKKLNKNDHILAITSGDNLVILYALTAKPARIHAVDMNPPCQGHILELK 487
 QY 336 KVAIQOLEPDPVWOLFEGVHPRIELEYEKKLAPFLSQTSHNFWSKRLWYFQHGLYVQGG 395
 Db 488 LAIAQALEYNDFWLIFEGEGRHPEDKLLITTKLSFFLS-----SCWALRL----- 531
 QY 396 MGKLCWVLQCLAVVLGLGKTVKRELANAPTWEEQRRLWDSNMLIHFKVNGPKPLVILFVKRF 455
 Db 532 -----AQIAFFIAGVRGDKELCOATSTAEQERIWQKKI-----RP-VFLNKVM 574
 QY 456 VSLVLFNKAVLWFGGVPKQYKQYALIKADGPIPIENYIARTWGDVAENSHVRKQNYFFYNCL 515
 Db 575 VKLFLGNLPLFNHVALGVPQNMNCFLDQG-SVEDYVKATLDPIPTLSTLKDNDYFFLLCL 633
 QY 516 TGKFLDNCPTLYLREAAFAFLK-SGVVDNLTVSTNFMELKA---RTVTKVILMDHVDW 571
 Db 634 NGRYTTSQPAFLKPGFKALNSKSTDAFKLHTDILNVLRLGRLPESLTKIIVMSMDW 693
 QY 572 LD-----MPVAN-----ELAECLAKQVAPGGIVIVRSAS 600
 Db 694 FDPPIPGTLPQGNSTALDTLOATPEKALEHLRAELDYELB-MKRVLVKGGIIVRSAA 752
 QY 601 LSPYLAELQKAGFVRCIR-RATQGYMDRVNMYSSFYMA 639
 Db 753 KRPWYRQFEAAGLVQPIRENOQAIDRVNMYASFKA 792

RESULT 12
 Q4X175 ASPFU PRELIMINARY; PRT; 831 AA.
 ID Q4X175; Q4X175; Q4X175; Q4X175; Q4X175; Q4X175; Q4X175; Q4X175; Q4X175; Q4X175;
 AC Q4X175; Q4X175; Q4X175; Q4X175; Q4X175; Q4X175; Q4X175; Q4X175; Q4X175; Q4X175;
 DT 13-SEP-2005 (T-EMBLrel. 31, Created)
 DT 13-SEP-2005 (T-EMBLrel. 31, Last sequence update)
 DT 13-SEP-2005 (T-EMBLrel. 31, Last annotation update)
 DE Hypothetical protein.
 GN ORFNames=Afu2g10950;
 OS Aspergillus fumigatus Af293.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
 OX NCBI_TaxID=330879;
 RN [1]
 RC NUCLEOTIDE SEQUENCE.
 RP STRAIN=Af293;
 RA Nierman W., Pain A., Anderson M.J., Wortman J., Kim H.Stanley.,
 RA Arroya J., Berriman M., Abe K., Archer D.B., Bermejo C., Bennett J.,
 RA Bowyer P., Chen D., Collins M., Coulson R., Davies R., Dyer P.S.,
 RA Farman M., Fedorova N., Fedorova N., Feldblyum T.V., Fischer R.,
 RA Foster N., Fraser A., Garcia J.L., Garcia M.J., Goble A.,
 RA Goldman G.H., Gomi K., Griffith-Jones S., Gwilliam R., Haas B.,
 RA Haas H., Harris D., Horiuchi H., Huang J., Humphrey S., Jimenez J.,
 RA Keller N., Khouri H., Kitamoto K., Kobayashi T., Kulkarni R.,
 RA Kumagai T., Latton A., Latge J.-P., Li W., Lord A., Lu C.,
 RA Majoros W.H., May G.S., Miller B.L., Mohamoud Y., Molina M., Monod M.,
 RA Mouyna I., Mulligan S., Murphy L., O'Neil S., Paulsen I.,
 RA Penava M.A., Perlea M., Price C., Pritchard B.L., Quail M.A.,
 RA Rabinowitz E., Rawlins N., Rajandream M.-A., Reichard U.,
 RA Renauld H., Robson G.D., Rodriguez de Cordoba S., Rodriguez-Pena J.M.,
 RA Ronning C.M., Rutter S., Salzberg S.L., Sanchez M.,
 RA Sanchez-Ferrero J.C., Saunders D., Seeger K., Squares R., Squares S.,
 RA Takeuchi M., Tekaita F., Turner G., Vazquez de Aldana C.R., Weidman J.,
 RA White O., Woodward J., Yu J.-H., Fraser C., Galagan J.E., Asai K.,
 RA Machida M., Hall N., Barrell B., Denning D.W.,
 RA "Genomic sequence of the pathogenic and allergenic filamentous fungus
 RT Aspergillus fumigatus.";
 RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
 CC -!- CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 DR EMBL; AAHF01000001; EAL93390.1; -; Genomic_DNA.
 DR InterPro; IPR000051; SAM_bd.

KW Hypothetical protein; Methyltransferase; Transferase.
 SQ SEQUENCE 831 AA; 94737 MW; D3601019DB480CB4 CRC64;
 Query Match 23.3%; Score 806.5; DB 2; Length 831;
 Best Local Similarity 29.7%; Pred. No. 8.8e-57;
 Matches 224; Conservative 95; Mismatches 257; Indels 177; Gaps 20;
 QY 42 KKG-DDHAARLSPFCYQQAAP-----AARL-----AERSNLIW 74
 Db 69 KKGALQDQDALESFYKTOAGVYDATRKRLCGREDMLGLVAAQLKYKVENKELQAGKAIW 128
 QY 75 VDLGGGTGNVDMADYIDLAKFKS-IYVVDLCHSLCEVAKKAKAGKWNQVWEADAC 133
 Db 129 VD-----DSNIEAMASFLPVTQFFSHVYLVDLSPLCEVARQERFELGKWNVTVCODAR 184
 QY 134 QF-----APEGTATLITFSYSLTWIPPHNVIDQACSYLSODGLGVGADFVY 181
 Db 185 SFRLPHEKVDPRAPAPSTAGADLITMSYLSLMPDYVSVDSLTDLRAGSILGVCDVYV 244
 QY 182 SKYDLP-----ROMPWRFRFWSIFDIDNIDIGPERRAYLEQKLERWEQNTQ 232
 Db 245 QSTVDVSARNYIGGAFNRHVNMLGRAFWRAFADRVSLAARRDYLFYRFGTVISASER 304
 QY 233 ----GSIYPVWLRAPYVWIGR-----251
 Db 305 NYLLGGI-----PYIIFVGRHKNLASLQGETIEMLDASFTSPYLSPANHTKEME 356
 QY 252 ----LPSVGHALHEERV-----ERPPMPPTFLYQ 278
 Db 357 KAMERNAQIRSKAYESAVINLSANLPLPSSFFQNHHCRIFYNDLLPKHTQFKNYIYAF 416
 QY 279 SWEDPEPDMVMEINPKDTVLTSLTSGCCNALNLLVQAGQGVSVDCNPAQSALLELKVA 338
 Db 417 NWEDPRVDHRLDIDKDDVILAITSGDNLVYLSQSPRRVHVDLNPQNHLLELKVAS 476
 QY 339 IQOLEPDPVWOLFEGVHPRIELEYEKKLAPFLSQTSHNFWSKRLWYFQ-----HGLYVQ 394
 Db 477 FMALGHRDVKIKFEGEGKHPFRELILSLUSAHLSSQAFQWLEHTHIFTSKYKGLYETG 536
 QY 395 GNGKLCWVLQCLAVVLGLGKTVKRELANAPTWEEQRRLWDSNMLIHFKVNGPKPLVLFVK 454
 Db 537 GSRHAIKMVRVYLFKVGSLGSKVCLCEAQTAEQREIWPKIRAVLM-----SKELHW--- 588
 QY 455 FVSLVFNKAVLWFGGVPKQYALIKAD-----GPIENYIARTMD 496
 Db 589 ---AVSTEFWFAKAGVPRNQNMIVDDYFKRLGLTKDMNQKDISGRSIWQYVVDTL 645
 QY 497 GVAENSHVRKQNYFYNCITGKFLRDNCTYLREAFATLKS-GVVDNLTVSTNFFWEEL 555
 Db 646 PVVNETMISNDNYFYFLCLOQFSRRCHPTYLSPOAHVKLSSPGAFDGLRIHTDEINEVI 705
 QY 556 K---ARTYTKVILMDHVDMDMPVANEL---AECLAKQVAPGGIVIVRSASISPPYVAELI 609
 Db 706 KRITPSLTIIVMSMDWFD-PEGTEASIQAKLHAKMDGRILLRSASIEPWIQKF 764
 QY 610 QKAGFDVRCI-RRATQGYMDRVNMYSSFYMAR 641
 Db 765 EENGFTARRVGARFPGSCIDRVNMYASTWICTK 797

RESULT 13
 Q5B0L6 EMBL PRELIMINARY; PRT; 790 AA.
 ID Q5B0L6; Q5B0L6; Q5B0L6; Q5B0L6; Q5B0L6; Q5B0L6; Q5B0L6; Q5B0L6; Q5B0L6; Q5B0L6;
 AC Q5B0L6; Q5B0L6; Q5B0L6; Q5B0L6; Q5B0L6; Q5B0L6; Q5B0L6; Q5B0L6; Q5B0L6; Q5B0L6;
 DT 10-MAY-2005 (T-EMBLrel. 30, Created)
 DT 10-MAY-2005 (T-EMBLrel. 30, Last sequence update)
 DE Hypothetical protein.
 GN ORFNames=AN5914.2;
 OS Aspergillus nidulans FGSC A4.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eurotiales; Trichocomaceae; Emericella.
 OX NCBI_TaxID=227321;

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: March 14, 2006, 01:53:31 ; Search time 104 Seconds
(without alignments)
2737.668 Million cell updates/sec

Title: US-10-620-914-45
Perfect score: 3463
Sequence: 1 MSGGRDPASVTKNFKSLE.....RVNMYSSFYMKRKGAKXN 648

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21.*
1: Geneseq1980s:*
2: Geneseq1990s:*
3: Geneseq2000s:*
4: Geneseq2001s:*
5: Geneseq2002s:*
6: Geneseq2003as:*
7: Geneseq2003bs:*
8: Geneseq2004s:*
9: Geneseq2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3463	100.0	648	9	Adw72748 Chlamydom
2	891	25.7	752	8	Adp98846 C. albica
3	850.5	24.6	908	9	Adw72753 Neurospor
4	842.5	24.3	845	8	Adr86056 Aspergill
5	381.5	11.0	416	9	Adw72756 Rhodobact
6	379.5	11.0	416	6	Abg72176 Rhodobact
7	379.5	11.0	416	9	Adw72706
8	326.5	9.4	416	6	Abg72178 Agrobacte
9	326.5	9.4	416	9	Adw72732 Agrobacte
10	321	9.3	415	9	Adw72744
11	292.5	8.4	416	6	Abg72180 Sinorhizo
12	292.5	8.4	416	9	Adw72736 Sinorhizo
13	123	3.6	2284	8	Adt04703 House mou
14	121	3.5	212	7	Adc97627 E. faeciu
15	120	3.5	221	6	Abg72181 Sinorhizo
16	120	3.5	221	9	Adw72738 Sinorhizo
17	119	3.4	2286	8	Adt04699 Human DNA
18	118	3.4	329	9	Abm97161 M. xanthu
19	112.5	3.2	672	6	Abu40254 Protein e
20	111	3.2	195	8	Adk48246 Streptoco
21	111	3.2	198	8	Adr95987 Novel S.
22	111	3.2	198	9	Aea59857 Streptoco
23	110.5	3.2	298	4	Abg07975 Novel hum
24	108.5	3.1	706	7	Ab076020 Pseudomon

ALIGNMENTS

RESULT 1

ID	ADW72748	standard; protein; 648 AA.
XX	ADW72748;	
AC	ADW72748;	
XX	ADW72748;	
DT	21-APR-2005 (first entry)	
XX	Chlamydomonas reinhardtii Btal gene, protein.	
DE	Chlamydomonas reinhardtii Btal gene, protein.	
XX	Phospholipid synthesis; transgenic plant; fertilizer; enzyme;	
KW	betaine lipid.	
XX	Chlamydomonas reinhardtii.	
OS	Chlamydomonas reinhardtii.	
XX	WO2005009115-A2.	
PN	03-FEB-2005.	
PD	15-JUL-2004; 2004WO-US022789.	
XX	16-JUL-2003; 2003US-00620914.	
XX	(UNMS) UNIV MICHIGAN STATE.	
PA	Benning C, Riekhof W;	
PI	WPI; 2005-112975/12.	
DR	N-PSDB; ADW72746, ADW72747.	
DR	New composition comprising a purified DNA having an oligonucleotide	
PT	sequence encoding a protein, useful in producing Betaine lipid compounds	
PT	e.g., Diacylglycerol-O-4'-(N,N,N-trimethyl) homoserine (DGTS).	
XX	Claim 11; SEQ ID NO 45; 147pp; English.	
PS	The invention relates to a new composition comprising a purified DNA	
XX	having an oligonucleotide sequence appearing as ADW72747 (Chlamydomonas	
CC	reinhardtii Btal cDNA) or ADW72752 (Neurospora crassa Btal coding region,	
CC	encoding Btal proteins which are DGTS (diacylglycerol-O-4'-(N,N,N-	
CC	trimethyl)homoserine) synthetic enzymes. Btal (and BtaA and BtaB) enzymes	
CC	allow the replacement of phospholipids with non-phosphorus containing	
CC	lipids in the cell membrane. A transgenic plant expressing the enzymes	
CC	would have a reduced need for phosphate-containing fertilizer. Also	
CC	included are an RNA transcribed from the purified DNA, antibodies	
CC	produced from the protein, a vector comprising the DNA or nucleic acid, a	
CC	host cell comprising the vector, transgenic plants comprising the vector,	

or for preventing or inhibiting formation on a surface of a biofilm comprising *Candida albicans*; a pharmaceutical composition comprising a therapeutic amount of an agent which reduces the activity or level of a gene product encoded by a nucleic acid comprising any of ADP98516-ADP98825 in a pharmaceutical carrier; an antibody preparation which binds the polypeptide; methods for evaluating a compound against a target gene product encoded by any of ADP98516-ADP98825; identifying an antimycotic compound; a computer or a computer readable medium that comprises at least one of the nucleotide sequences mentioned in the specification or at least one amino acid sequence selected from ADP98826-ADP99135; a method assisted by a computer for identifying a putatively essential gene of a fungus; and a protein array comprising proteins, where at least one protein comprises an amino acid sequence or a portion of an amino acid sequence selected from ADP98516-ADP98825. The novel methods and compositions have fungicide activity. The compositions may be used in gene therapy. The composition and methods are useful for drug screening purposes or for diagnosing, preventing or treating infections associated with *Candida albicans*. These may also be used for constructing strains useful for identification and validation of gene products as effective targets for therapeutic intervention, for identifying and validating gene products as effective targets for therapeutic intervention, and for collecting identified essential genes. This sequence represents the protein of a *Candida albicans* fungal specific gene of the invention. CC NOTE: This sequence was downloaded from an electronic sequence listing provided on the WIPO website. CC XX SQ Sequence 752 AA;

Query Match 25.7%; Score 891; DB 8; Length 752;
Best Local Similarity 32.7%; Pred. No. 1.2e-82;
Matches 229; Conservative 98; Mismatches 232; Indels 142; Gaps 21;

QY 51 LESFYGFQAAAF-----AARLAERSNLIWDLGGGTGENVDMADVI 92
DB 70 LESFYQNQAHYDNTREPLKGRQECRLRAISHLPKKKLLIWDICGGTSGNIEFDEIS 129
QY 93 DLAK-FKSIYVDLCHSLCEVAKKAKAKGKQVVEADACQAPPEGTATLITFSYL 151
DB 130 KISENFKAVYLDLSPSLCEVAKARPEAHWTNVHVLVADACDFTIDYSDADLITFSYL 189
QY 152 TMIPPHNVIDQACVLSQDLGVADFYVS-----GKYD-----LPLQMPWSRRFFWR 201
DB 190 SMIPFTNAIDNAVSKLDMEGHIIATDFGICSSDTSMTGRINTVGGVLRNDRIDPILNFWR 249
QY 202 SIFDIDNIDIGPERRAYLEOKLERVWEQNTQ-----GSIPVPLWLRAPYVYVIGRLPSVGH 257
DB 250 IYFEADKVLDSRRRYLEVYKFTVKLSNYSNKALGI-----PYTIWIGCDKSKSH 301
QY 258 ALHEE-----RVERPPMPPT-----273
DB 302 TILERLNLCLATESPYLAPTTPITIANQLEDIPISKGHEAALINLQKLPVPSMYQKEYWR 361
QY 274 -----FLYQSWEDPEPDMVMEINPKDTVLTLTSGGCVNALML---LVQG 315
DB 362 VYVDENMPLYEQPKQYIYFTWEDPREHKLNFSTDDTTLAITSAGDNILSYASLPTP 421
QY 316 AGOVSVDCNPAQSALLEKVAIQLEPERVWOLFEGVHPRIELEYKELAPFLSQTS 375
DB 422 PKKIHAVDLNPQNHLLEKLASFRCLSQEQIWSMFGEGKIEFNFDLLIDTLAPHMSSNA 481
QY 376 HNFWSKR--LWYFQHLGYTQGGMGKLCWLQC---LAVVLGLGKTVKRLANAPTMEEQRR 430
DB 482 FQYMDMDKPKTFSGKGLY---DTGFSRMAWLRSLRYFKVCGVSKYVEELCAATTMEEQLR 538
QY 431 LWDNSMLIHVRKNGPKPLVWLFVKFV-SLVLFNKAVLWFGGVGPKQYKQYALIKADGPIEN 489
DB 539 INWEHL-----KPT--LFPVPVGSLLVGNPFLWKLALGVPAQAALM---GPSVIK 584
QY 490 YIARTWDGVAENSHVRKQNYFYNCVLTGKFLRDNCTYLRERAAATLK-----SGVV 541
DB 585 YVVDLTLDPIIKRSMISNDNYFYFLCMGMYRTKNCNPDYLTITTKGFNRLSSTAATASGSSPI 644
QY 542 DNLTVSTNPFME---ELKARTYTKVILMDHVDWLD---MPVANELAECLAKQVAPGGIVI 595

DB 645 DNLRIHTDITLNEVFGRLKEKSIITAIIMDMDFDNGRDAINEIT-ALKRCLAPGGRVL 703
QY 596 WRSASLSPPYAEILQKAGF-DVRCIRRATQGYMDRVNMYSS 635
DB 704 LRSASTKPYLTKTFKMLGFGQSEENVVRQPGSSIDRVNMYAN 744

RESULT 3
ADW72753
ID ADW72753 standard; protein; 908 AA.
XX
AC ADW72753;
XX
DT 21-APR-2005 (first entry)
XX
DE *Neurospora crassa* Btal gene, protein.
XX
KW Phospholipid synthesis; transgenic plant; fertilizer; enzyme; betaine lipid.
XX
OS *Neurospora crassa*.
XX
PN WO2005009115-A2.
XX
PD 03-FEB-2005.
XX
PF 15-JUL-2004; 2004WO-US022789.
XX
PR 16-JUL-2003; 2003US-00620914.
XX
PA (UNMS) UNIV MICHIGAN STATE.
XX
PI Benning C, Riekhof W;
XX
DR WPI; 2005-112975/12.
XX
N-PSDB; ADW72751, ADW72752.
XX
PT New composition comprising a purified DNA having an oligonucleotide sequence encoding a protein, useful in producing Betaine lipid compounds e.g., Diacylglycerol-O-4'-(N,N,N-trimethyl) homoserine (DGTS).
XX
PS Claim 11; SEQ ID NO 50; 147pp; English.
XX
CC The invention relates to a new composition comprising a purified DNA having an oligonucleotide sequence appearing as ADW72747 (*Chlamydomonas reinhardtii* Btal cDNA) or ADW72752 (*Neurospora crassa* Btal coding region, encoding Btal proteins which are DGTS (diacylglycerol-O-4'-(N,N,N-trimethyl)homoserine) synthetic enzymes. Btal (and BtaA and BtaB) enzymes allow the replacement of phospholipids with non-phosphorus containing lipids in the cell membrane. A transgenic plant expressing the enzymes would have a reduced need for phosphate-containing fertilizer. Also included are an RNA transcribed from the purified DNA, antibodies produced from the protein, a vector comprising the DNA or nucleic acid, host cell comprising the vector, transgenic plants comprising the vector, and a protein translated from the RNA or encoded by the nucleic acid. The composition is useful in producing Betaine lipid compounds e.g., Diacylglycerol-O-4'-(N,N,N-trimethyl) homoserine (DGTS). The composition of the invention reduces the amount of phosphate fertilizer needed for the optimal growth of crop plants. The present sequence represents a betaine lipid synthetic enzyme.
XX
SQ Sequence 908 AA;

Query Match 24.6%; Score 850.5; DB 9; Length 908;
Best Local Similarity 30.4%; Pred. No. 2.8e-78;
Matches 222; Conservative 105; Mismatches 235; Indels 169; Gaps 18;

QY 69 RSNLIWDLGGGTGENVDMADYIDLAK-FKSIYVVDLCHSLCEVAKKAKAKGKQV 127
DB 178 RKKPIWVDVGGGTGNIEAWKAPVNVSEFKTIVLDFPSLCEVAKPARLGNVRV 237
QY 128 VEADACQFA-----PPEGT-----ATLITSYSLTIPPF 157

Db 238 ICTDARKFLREDYEDYDEGSGSDSPSLSGWGETKPGRHAGALITMSYSLWMPDY 297
Qy 158 HNVIDQACSVLSODGLVGVADVFVSGKYD-----LPLROMPSRRFFWRSIFDIDN 208
Db 298 FSIIDSLESLLAPHLIAVDFVFAQSKVDFTFRNYTGGLMNRHVGFARNFWSWFDADR 357
Qy 209 IDIGPERRAYLEOKLERVW-----EONTQSGIPYVPW---LRAPYVW----- 247
Db 358 VSLPARROVLEVRFGVTLVNARNNTLGAIPYIWLGLCKKFPSTSSLPHEIVEHIDAI 417
Qy 248 -----WIGR-----LPSVGHALH 260
Db 418 ATEPSRSSPLVKGHSSSATNALAFVAGRTAPEMRSKAFNTAIENISANLPLSPFFYQNH 477
Qy 261 -----EERVERPMPPTFLYQSWEDPEPDMEVMEINPKDVTLTITSGGCNALNLVQ 314
Db 478 HWRYYDDQLPKHTQNDENYIYFTWEDSVRELLNLGDDVVLTAITSAGDNLISYLMQ 537
Qy 315 GAGQVSVDCNPAQSALLELKKVAIQOLEPEDVWQVLFGEVHPRIEELYEKKLAPFLSQ 374
Db 538 SPARVHAIDLNPQNHLLELKVASFYTLDPVWKIFGEGKHPDFRSLILSKLSPHLSGR 597
Qy 375 SHNFWSKRLWYFO-----HGLYYQGGMGKLCWVLOCLAVVLGLGKTVKRLANAPTMBEQR 430
Db 598 AFQYWLNSNAHIFTDPAGRGVDTGGSRYAIRFRFWISTLFFCRSAVRRLSTPTLEGQRS 657
Qy 431 LWDNSMLIHFKVNGKPLWLVFKVYS-LVLFNKAVLWFGGVGPKQYALIKAD----- 483
Db 658 IYHTKI-----RPC--LLNRFFVNLVSSDAFLWSALGVFKQKQVAMIEADYHRSI 706
Qy 484 -----GPIENYIARTWDGVAENSHVRKQNYFYNCILTKGFLRNCPTYL 529
Db 707 SSTTSSKEKPSRAEAILHYTTSTLDPVLSHLSADNPYILVCLVQYTRQCHPDYLS 766
Qy 530 EAPATLKS-GVVDNLTVSNFMEEL---KARTYTKVILMDHVDMLDMP----- 575
Db 767 PAAHSILSAPGAFDGLRIHTDEIQEVLARFQPGTLTAVVMDSDMDFDPPSPBEEKEGR 826
Qy 576 VANELAECLAQVAPGIVTWRSASLSPPYAELIQKAGFDVRCI-----RATQGYMDRV 630
Db 827 KAREQVRLNRALKVGGKVLRSAGVEPWYRVFVEEGFARRVGRESGRGQBCIDRV 886
Qy 631 NMYSSFFYMAR 641
Db 887 NMYASWILEK 897

RESULT 4
ID ADR86056
XX ADR86056 standard; protein; 845 AA.
AC ADR86056;
XX
DT 04-NOV-2004 (first entry)
XX
DE Aspergillus fumigatus essential gene protein #106.
XX
KW Fungicide; Aspergillus fumigatus infection; Farmer's lung disease;
KW drug screening.
XX
OS Aspergillus fumigatus.
XX
PN W02004067709-A2.
XX
XX 12-AUG-2004.
XX
XX 16-JAN-2004; 2004WO-US001099.
XX
XX 17-JAN-2003; 2003US-0441281P.
XX
XX 13-JUN-2003; 2003US-0478196P.
XX
PA (ELIT-) ELITRA PHARM INC.

(ELIT-) ELITRA CANADA LTD.
XX Jiang B, Hu W, Lemieux S, Roemer T;
XX WPI: 2004-594200/57.
XX N-PSDB; ADR85469.
XX
XX New purified or isolated Aspergillus fumigatus nucleic acid molecule
XX encoding a gene product, useful for diagnosing and/or treating invasive
XX fungal infections, such as Farmer's lung disease.
XX
XX Claim 1; SEQ ID NO 3106; 164pp; English.
XX
XX The present invention relates to Aspergillus fumigatus genes that are
XX essential and are potential targets for drug screening. The methods and
XX compositions of the present invention are useful for diagnosing and/or
XX treating invasive Aspergillus fumigatus infection, including the allergic
XX forms of the disease, such as Farmer's lung disease. They can also be
XX used in various drug discovery purposes, such as expression of the
XX recombinant protein, hybridization assay and construction of nucleic
XX arrays. The present sequence represents an Aspergillus fumigatus
XX essential gene protein sequence, used during diagnosis and drug
XX development in the invention. These genes share a high degree of sequence
XX conservation with known essential genes of candida albicans. The sequence
XX data for this patent is not represented in the printed specification, but
XX was obtained in electronic format from WIPO.
XX
XX SQ Sequence 845 AA;
XX
XX Query Match 24.3%; Score 842.5; DB 8; Length 845;
XX Best Local Similarity 30.1%; Pred. No. 1.7e-77;
XX Matches 230; Conservative 96; Mismatches 254; Indels 183; Gaps 21;
XX
Qy 42 KKG-DDHAARLESFYGPQAAAF-----AARL-----AERSNLIW 74
Db 69 KGLNCQQDALESFYTKQGVDTAKRLCGEDMLGLVAAQLKVKENKELQAGKAIW 128
Qy 75 VDLGGGTGENDVMADYIDLAKFS-IYVVDLCHSLCEVAKKAKAKGKNQVQVBRADAC 133
Db 129 VDIGGGTGYNEAMASFLPTQPFHVLVLDLSPSLCEVARQFERLGNVWVVCQAR 188
Qy 134 QF-----APPEGTATLITSYSLTMTIPPHNVIDQACSVLSQDGLGVADFYV 181
Db 189 SFRLPHEKVDPRAPAPSTAGADLITMSYSLMIDYYSVVDLTLDRASGLGVCDFYV 248
Qy 182 SGKYDLPL-----ROMPSRRFFRSIFDIDNIDIGPERRAYLEOKLERVWEQNTQ 232
Db 249 QSIVDVSARNYTGAFNRHVNLGRAFWADRVSLAEARDVLEYRFGVVISASER 308
Qy 233 ---GSIPYVPLRAPYVWIGR-----LPSVGHALHERV-----ERPMPFPFELYTQ 278
Db 309 NYLLGGI-----FYIFVGRHKNLASNLGQETIEMLDASFTSPYLSPANHTKEME 360
Qy 252 -----LPSVGHALHERV-----ERPMPFPFELYTQ 278
Db 361 KAMERNAQETRSKAYESAVINLSANLPLSPSFYQNHHCRIFYNDLLPKTKFKNYIYAF 420
Qy 279 SWEDPEPDMEVMEINPKDVTLTITSGGCNALNLVQAGOVSVDCNPAQSALLELKKVA 338
Db 421 NWEDPRVDHRLDIDKDDVLAITSAGDNLIDYLOKSPRVHADVLDNPNQNHLELKVAS 480
Qy 339 IOOLEPEDVWQVLFGEVHPRIEELYEKKLAPFLSOTSHNFWSKRLWYFO-----HGLYYQ 394
Db 481 FMAIGHRDVWKIFGEGKHPFEELLISRLSAHLSLSSQAFQWLEBHTHTFTSKYKGKLYETG 540
Qy 395 GMKGLWVLOCLAVLGLGKTVKRLANAPTMBEQRRLWDSNMLIHFKVNGKPLVWLVFK 454
Db 541 GSRHAIKMVRYLEFKVFGLEGQVKLCCEAQTLABQREIWPKIRAVLM-----SKPLHW----- 592
Qy 455 FVSLVLFNKAVLWFGGVGPKQYALIKAD-----GPIENYIARTMD 496
Db 593 ---AVSTEWFAWKAAGVPRNQRNMIVDDYFKLGLTKDMNQGDIGRSIWIQVVDLTD 649

QY 497 GVAENSHVRKQNYFYNYCLTGKFLR-----DNC-----PTYLREAPATILKS-GVVDNLT 545
 Db 650 PVNYNETMISNDNYFYCLQGSRRVQTNQSKCIRDXTYLSQARHVKLSSPGAFDGLR 709
 QY 546 VSTNPFMEELK---ARTYTKVILMDHVDLMDMPVANEL---ASCLAKQVAPGGIVWRSA 599
 Db 710 IHTDEINEVKRITPSLTITAVIMDSKWFDP-PEGTEASTQAKLNHALKMDGRILLRSA 768
 QY 600 SLSPPYAELIQKAGFVRCI-RRATQGYMDRVNMYSSFYNNARR 641
 Db 769 SIEPFWYIKQPEENGFTARRVGARFPGSGCIDRVNMYASTWICTK 811

RESULT 5

ADW72756
 ID ADW72756 standard; protein; 416 AA.

XX AC ADW72756;
 XX DT 21-APR-2005 (first entry)
 XX Rhodobacter sphaeroideis btaA gene, protein, L91 mutant.
 DE Phospholipid synthesis; transgenic plant; fertilizer; enzyme;
 KW betaine lipid; muten.
 KW Rhodobacter sphaeroideis.
 OS Synthetic.
 XX WO2005009115-A2.
 XX 03-FEB-2005.
 XX 15-JUL-2004; 2004WO-US022789.
 XX 16-JUL-2003; 2003US-00620914.
 XX (UNMS) UNIV MICHIGAN STATE.
 XX Benning C, Riekhof W;
 XX WPI; 2005-112975/12.

PT New composition comprising a purified DNA having an oligonucleotide
 PT sequence encoding a protein, useful in producing Betaine lipid compounds
 PT e.g., Diacylglycerol-0-4'-(N,N,N,-trimethyl) homoserine (DGTS).

PS Example 3; Page; 147pp; English.

CC The invention relates to a new composition comprising a purified DNA
 CC having an oligonucleotide sequence appearing as ADW72747 (Chlamydomonas
 CC reinhardtii Bta1 cDNA) or ADW72752 (Neurospora crassa Bta1 coding region,
 CC encoding Bta1 proteins which are DGTS (diacylglycerol-0-4'-(N,N,N,-
 CC trimethyl)homoserine) synthetic enzymes. Bta1 (and BtaA and BtaB) enzymes
 CC allow the replacement of phospholipids with non-phosphorus containing
 CC lipids in the cell membrane. A transgenic plant expressing the enzymes
 CC would have a reduced need for phosphate-containing fertilizer. Also
 CC included are an RNA transcribed from the purified DNA, antibodies
 CC produced from the protein, a vector comprising the DNA or nucleic acid, a
 CC host cell comprising the vector, transgenic plants comprising the vector,
 CC and a protein translated from the RNA or encoded by the nucleic acid. The
 CC composition is useful in producing Betaine lipid compounds e.g.,
 CC Diacylglycerol-0-4'-(N,N,N,-trimethyl) homoserine (DGTS). The composition
 CC of the invention reduces the amount of phosphate fertilizer needed for
 CC the optimal growth of crop plants. The present sequence represents a
 CC mutated R. sphaeroideis BtaA or BtaB protein. NOTE: The present sequence
 CC is not shown in the specification but was created by the indexer using
 CC the wild-type protein and the information in example 3. Only one of the 3
 CC BtaA mutants could be created, the other two (A201G and S399T) referred
 CC to amino acids not present at those positions in the wild-type sequence.

XX Sequence 416 AA;

Query Match 11.0%; Score 381.5; DB 9; Length 416;
 Best Local Similarity 28.1%; Pred. No. 5.5e-30;
 Matches 127; Conservative 68; Mismatches 182; Indels 75; Gaps 20;
 QY 231 TQGSIVYVWLRAPYVYVWIGRLPSVGHALH-----EERVERPPMPFPPTF---LVTQSW 280
 Db 2 TQFALTHIP---APP---VAR---QIGAAVHRTSLLSAEGLMER--MPSRLFHGLVYVQIW 51
 QY 281 EDPEPDMYMEINPKDVTVLTLTSGGCNVALNLLVQAGQVSVVDCNPAQSALLBLKKVAIQ 340
 Db 52 EDPAVDMAALAIRPGDRLVAIASGGCNVLSYLTQGPGLAVDLSPAHVALGRKLKLAAR 111
 QY 341 QL-EPEDVWQLPGEVHPRIEELYEKKLAPFLSQTSHNF-----SKRLWYFQHGGLYYQ 393
 Db 112 TLPDHAADFLLFGRADLPNGAALYDRHIAIPALDGRSRRYWEARSPPGRRITQLFERGFYRH 171
 QY 394 GGMGLCWVLQCLAVVLGLGKTVKRLANAPTMBEQRRLWDSNMLIHVKNQPKPLVWLTV 453
 Db 172 GALGRFIGAHTLA--RAAGTDLRGFLDCPDTEAQRSFFVAHI-----GP-----LPE 217
 QY 454 KFVSLVLFNKAVLWFGGVPKGQYALIKADG----IPIENYIARTMDGVAENSH-----503
 Db 218 APVQALARRPAALFGLGIPPAQYALLAGDGDGVLPV-----LRQRLHRLLCDF 267
 QY 504 VRKQNYFYNYCLTGKFLRN---CPTYLREAPATILKSGVVDNLTVTSTFFMEELKARTY 560
 Db 268 PLRENYFAQAIARRYPFEGEGALPPYLEPTAFETLREN-AGRVQIENRSLTALAARPE 326
 QY 561 TKV---ILMDHVDLMDMPVANELACLAKQVAPGGIVWR---SASISPPYAEILQKAGF 614
 Db 327 ESIHGFTLLDAQDMTDAQLTALWQVTRTAAPGARVIFRTGGAULLP--GRVPEILG 384
 QY 615 DVRCIRRTAQG--YMDRVNMYSSFYNNARRKA 644
 Db 385 HWRADRAAGQAGHAADRSALYGGFHLVRRDA 416

RESULT 6

ABG72176
 ID ABG72176 standard; protein; 416 AA.

XX AC ABG72176;
 XX DT 05-FEB-2003 (first entry)
 XX Rhodobacter sphaeroideis btaA protein.
 DE Betaine lipid production; betaine lipid compound; DGTS;
 KW diacylglycerol-0-4'-(N,N,N,-trimethyl) homoserine; agricultural industry;
 KW phosphate-containing fertilizer; btaA; enzyme.
 XX Rhodobacter sphaeroideis.
 PH Key Location/Qualifiers
 FT Misc-difference 1 /note= "Encoded by GTC"

XX WO200283844-A2.

XX 24-OCT-2002.

XX 09-APR-2002; 2002WO-US011134.

XX 13-APR-2001; 2001US-0283812P.

XX 08-APR-2002; 2002US-00283812.

XX (UNMS) UNIV MICHIGAN STATE.

XX Benning C, Riekhof W, Klug R;

XX WPI; 2003-058632/05.

XX N-PSDB; ABS58051.

XX


```

Db 327 ESINGFTLLDAQDWMTDALQTLWRQVTRTAAPCARVIFRTGGAADLP--GRVPEILG 384
QY 615 DVCCIRATOG--YMDRVNMYSSFYMARRKA 644
Db 385 HWRADRAAGQAGHAADRSAYGGFHLVRRDA 416

RESULT 8
ABG72178
ID ABG72178 standard; protein; 416 AA.
XX
AC ABG72178;
DT
DT 05-FEB-2003 (first entry)
XX
DE Agrobacterium tumefaciens btaA protein.
XX
KW Betaine lipid production; betaine lipid compound; DGTS;
KW diacylglycerol-0-4'-(N,N,N,-trimethyl) homoserine; agricultural industry;
KW phosphate-containing fertilizer; transgenic; btaA; enzyme.
XX
OS Agrobacterium tumefaciens.
XX
PN WO200283844-A2.
XX
PD
PD 24-OCT-2002.
XX
PF 09-APR-2002; 2002WO-US011134.
XX
PR 13-APR-2001; 2001US-0283812P.
PR 08-APR-2002; 2002US-00283812.
XX
PA (UNMS ) UNIV MICHIGAN STATE.
XX
PI Benning C, Riekhof W, Klug R;
XX
XX WPI; 2003-058632/05.
DR N-PSDB; ABS58082.
XX
XX New composition comprising an isolated and purified DNA molecule, useful
PT for producing betaine lipids, e.g. Diacylglycerol-0-4'-(N,N,N,-trimethyl)
PT homoserine (DGTS) for agricultural applications.
XX
PS Claim 11; Fig 25; 109pp; English.
XX
CC The present invention relates to compositions and methods for producing
CC betaine lipids. The method comprises the expression of recombinant
CC enzymes (e.g. from Rhodospirillum rubrum) in host cells such as
CC bacteria, yeast, and plants to produce betaine lipid compounds including
CC diacylglycerol-0-4'-(N,N,N,-trimethyl) homoserine (DGTS). The methods and
CC compositions of the invention are useful in agricultural applications,
CC such that the amount of phosphate-containing fertilizer required for the
CC growth of a particular plant is decreased. The polynucleotide sequences
CC encoding the recombinant enzymes may be used to produce vectors which can
CC be used to produce transgenic plants. The present sequence represents the
CC R. sphaeroides btaA homologue from Agrobacterium tumefaciens
XX
XX Sequence 416 AA;

Query Match 9.4%; Score 326.5; DB 6; Length 416;
Best Local Similarity 27.8%; Pred. No. 2.9e-24;
Matches 111; Conservative 69; Mismatches 158; Indels 61; Gaps 16;

QY 275 LYTQSWEDPDPDMEVMEINPKOTVLTITSGGCNMLLVQAGQVSVSDCNPAQSALLEL 334
Db 44 VYFQIWEDEPDEMEALGEGHRIIVITGSGGCNMLAYLSRNPASIDVDLNPHEALNKL 103
QY 335 KVAIQQL-EEFVDVQLFGGVHPIREELVEKPLAPLSTSHNFSK-----RLWVF 386
Db 104 KLAARHLPAHQDVRVHFGGAGRTSRNSVGVDRFIAEHLDDATKAYWSKRTLSGRRISVF 163
QY 387 QHGLYYGGGKGLCWVLQCLAVLGLGKTVKRLANAPTMBEQRRLLWDSNNLIHFVKNPK 446

Db 164 DRNIYRTGLGRFIGAGHIMARLHGVLKLT--ENAKTRTLDQRFDSKVAPLF-----DK 217
QY 447 PLV-WLFFKVSLSLVFNKAVLMFGGVPKQY---ALIKADGIPPIENYIARTWDGVAENS 502
Db 218 PVVRWLTKKSSL-----FGLGIPPRQVDELASLSDG-TVASVLKERLEKLACNF 267
QY 503 HVRKONYFYNCITGKFLRDN---CPTYLREAFATLKSGVDNLTVSTNFFMEELK--- 556
Db 268 PL-SDNYFAWQAFARYPPEPHEGALPAYLKPEYKIRNNTA-RVAVHHATYTELLSRKP 325
QY 557 ARTYTKVILMDHVDWLDMPVANELACLAKQVAPGGIVWRSAS-----LSPVPAEL 608
Db 326 ANGVDRIYILLDAQDWMTDVQLNELMSQISRTAASGARVIFRTAARKSVIEGRUSP----- 380
QY 609 IQKAGPDVR-----CIRRATQ-GYMDRVNMYSSFYMARR 641
Db 381 -----DIRNQWYLEERSNELNMDRSAYGGFHLVQR 413

RESULT 9
ADW72732
ID ADW72732 standard; protein; 416 AA.
XX
AC ADW72732;
XX
DT 21-APR-2005 (first entry)
XX
DE Agrobacterium tumefaciens btaA gene, protein.
XX
KW Phospholipid synthesis; transgenic plant; fertilizer; enzyme;
KW betaine lipid.
XX
OS Agrobacterium tumefaciens.
XX
PN WO2005009115-A2.
XX
PD 03-FEB-2005.
XX
PF 15-JUL-2004; 2004WO-US022789.
XX
PR 16-JUL-2003; 2003US-00620914.
XX
PA (UNMS ) UNIV MICHIGAN STATE.
XX
PI Benning C, Riekhof W;
XX
XX WPI; 2005-112975/12.
DR N-PSDB; ADW72731.
XX
XX New composition comprising a purified DNA having an oligonucleotide
PT sequence encoding a protein, useful in producing Betaine lipid compounds
PT e.g., Diacylglycerol-0-4'-(N,N,N,-trimethyl) homoserine (DGTS).
XX
PS Disclosure; SEQ ID NO 29; 147pp; English.
XX
CC The invention relates to a new composition comprising a purified DNA
CC having an oligonucleotide sequence appearing as ADW72747 (Chlamydomonas
CC reinhardtii Bta1 cDNA) or ADW72752 (Neurospora crassa Bta1 coding region,
CC encoding Bta1 proteins which are DGTS (diacylglycerol-0-4'-(N,N,N,-
CC trimethyl)homoserine) synthetic enzymes. Bta1 (and BtaA and BtaB) enzymes
CC allow the replacement of phospholipids with non-phosphorus containing
CC lipids in the cell membrane. A transgenic plant expressing the enzymes
CC would have a reduced need for phosphate-containing fertilizer. Also
CC included are an RNA transcribed from the purified DNA, antibodies
CC produced from the protein, a vector comprising the DNA or nucleic acid, a
CC host cell comprising the vector, transgenic plants comprising the vector,
CC and a protein translated from the RNA or encoded by the nucleic acid. The
CC composition is useful in producing Betaine lipid compounds e.g.,
CC Diacylglycerol-0-4'-(N,N,N,-trimethyl) homoserine (DGTS). The composition
CC of the invention reduces the amount of phosphate fertilizer needed for
CC the optimal growth of crop plants. The present sequence represents a
CC betaine lipid synthetic enzyme.

```

```

XX SQ Sequence 416 AA;
Query Match 9.4%; Score 326.5; DB 9; Length 416;
Best Local Similarity 27.8%; Pred. No. 2.9e-24;
Matches 111; Conservative 69; Mismatches 158; Indels 61; Gaps 16;

QY 275 LYTSQWEDPPEDMEVMEINPKDVTLTSTGGCNALMLLVQAGQVSVDCNPAQSALLEL 334
DB 44 VYQIWEDEPIDMEAMELGEGHRIIVIGSGCNWLYLSRNPASIDVVDLPHHIALNKL 103
QY 335 KKVAIQOL-EFEDWQOLFEGVHPRIEELYEKKLAPFLSOTSHNFSK-----RLMYF 386
DB 104 KLAAFRLPAHQDVRHFRAGTRSNVGVDRFIAEHLDTATKAYNSKRTLSGRRRISVF 163
QY 387 QHGLYYQGGMGKLCWLQCLAVLGLKTVKRLANAPTMEEQRLWDSNMLIHVFNKGP 446
DB 164 DRNIYRTGLLGRFIGAGHINARLHGKLT--EMAKRTLTDEQRFDSKVAPLF----DK 217
QY 447 PLV-WLPVKFVSLVLFNKAFLWFGGVPKQY---ALIKADGIPNIENYIARTMDGVAENS 502
DB 218 PVRWLTKRSSL-----FOLGIPPRQYDELASLSDG-TVASVLKERLEKLACNF 267
QY 503 HVRKQNYFYFYNCLTGKFLRN---CPTYLREAAFPATLKSGVDNLTVSTNFFMEELK--- 556
DB 268 PL-SDNYFAWQAFARRYPEPEHEGALPAYLKPEYVEKIRNNTA-RVAVHHATYTELLSRKP 325
QY 557 ARTYTKVILMDHVDLMDPFAVLAELAECLAKQVAPGGIVWRSAS-----LSPPYAEL 608
DB 326 ANGVDRIYLLDAQDWMQTDVQLNELWSQISRTAASGARVIFRTAAEKSVIEGRILSP---- 380
QY 609 IQKAGFDVR-----CIRRTAQ-GYMDRVNMYSSFFYMARR 641
DB 381 -----DIRNQWVLEERSNELNAMDRAIYGGFHHIYQR 413

RESULT 10
ADW72744
ID ADW72744 standard; protein; 415 AA.
XX
AC ADW72744;
XX
DT 21-APR-2005 (first entry)
XX
DE Mesorhizobium loti btaA gene, protein.
XX
KW Phospholipid synthesis; transgenic plant; fertilizer; enzyme;
KW betaine lipid.
XX
OS Mesorhizobium loti.
XX
PN WO2005009115-A2.
XX
PD 03-FEB-2005.
XX
PF 15-JUL-2004; 2004WO-US022789.
XX
PR 16-JUL-2003; 2003US-00620914.
XX
PA (UNMS ) UNIV MICHIGAN STATE.
XX
PI Benning C, Riekhof W;
XX
DR WPI; 2005-112975/12.
DR N-PSDB; ADW72725.
XX
XX New composition comprising a purified DNA having an oligonucleotide
PT sequence encoding a protein, useful in producing Betaine lipid compounds
PT e.g., Diacylglycerol-O-4'-(N,N,N-trimethyl) homoserine (DGTS).
XX
PS Disclosure; SEQ ID NO 41; 147pp; English.
XX
CC The invention relates to a new composition comprising a purified DNA

```

```

CC having an oligonucleotide sequence appearing as ADW72747 (Chlamydomonas
CC reinhardtii BtaI cDNA) or ADW72752 (Neurospora crassa BtaI coding region,
CC encoding BtaI proteins which are DGTS (diacylglycerol-O-4'-(N,N,N-
CC trimethyl)homoserine) synthetic enzymes. BtaI (and BtaA and BtaB) enzymes
CC allow the replacement of phospholipids with non-phosphorus containing
CC lipids in the cell membrane. A transgenic plant expressing the enzymes
CC would have a reduced need for phosphate-containing fertilizer. Also
CC included are an RNA transcribed from the purified DNA, antibodies
CC produced from the protein, a vector comprising the DNA or nucleic acid, a
CC host cell comprising the vector, transgenic plants comprising the vector,
CC and a protein translated from the RNA or encoded by the nucleic acid. The
CC composition is useful in producing betaine lipid compounds e.g.,
CC diacylglycerol-O-4'-(N,N,N-trimethyl) homoserine (DGTS). The composition
CC of the invention reduces the amount of phosphate fertilizer needed for
CC the optimal growth of crop plants. The present sequence represents a
CC betaine lipid synthetic enzyme.
XX
SQ Sequence 415 AA;
Query Match 9.3%; Score 321; DB 9; Length 415;
Best Local Similarity 26.2%; Pred. No. 1.1e-23;
Matches 113; Conservative 70; Mismatches 176; Indels 72; Gaps 18;

QY 255 VGHALHEERV-----ERPPMPPTPF---LYTSQWEDPPEDMEVMEINPKDVTLTSG 304
DB 16 VQKAVYQNRALS KAGISER--LFAFLFGLVYVQIWEDEPDVDMQALQGQGHRIVTIASG 73
QY 305 GCNALMLLVQAGQVSVDCNPAQSALLELKKVAIQOLEFE-DVWQLFGSVHPRIEELY 363
DB 74 GCNLAFLYLRSPARIDAVIDLNAHAIALNRMKLEAVRRLPSQGLFRFFFGAADTSHNSQAY 133
QY 364 EKKLAPFLSQTSHNFSKRLW-----YFQHGGLYYQGGMGKLCWLQCLAVVLG----- 411
DB 134 DRFIAHLDPVSHYWERNRWGRRIAVDRNFYQTGLLG-----LFIAMGHRTAK 185
QY 412 -LGKTVKRLANAPTMEEQRLWDSNMLIHVFNKGPPLVWLFVKVLSVLFNKAVLWFGG 470
DB 186 FFGVNPAMWMEARNIGEQRRFFNEELAPVFDK---KLLKWATSRKASL-----FGL 233
QY 471 GVPGKQY--ALIKADGIPNIENYIAR---TMDGVAENSHVRKONYFYFYNCLTGKFLRDN- 523
DB 234 GIPPAQYDSLITSGDGTMAVSKARLEKLACDPFLEN-----NYFAWQAFARRYPNPG 287
QY 524 --CPTYLREAAFPATLKSGVDNLTVSTNFFMEEL---KARTYTKVILMDHVDLMDPFAV 578
DB 288 AALPAYLEKQNYETIR-GNIDRVAIHANLIEFLAGKDACTVDRFILLDAQDWMQTDQLN 346
QY 579 ELAECLAKQVAPGGIVWRSAS---SLSP--PYAELIQKAGFDVRCIRRTATQGYMDRVNMY 633
DB 347 ALWSEISRATASAGARVIFRTAABPSLLPGRVSTSLDDQMDYQDEASREFSA--RDRSAIY 404
QY 634 SSFYMARCKGA 644
DB 405 GGFHYVYKRTA 415

RESULT 11
ABG72180
ID ABG72180 standard; protein; 416 AA.
XX
AC ABG72180;
XX
DT 05-FEB-2003 (first entry)
XX
DE Sinorhizobium meliloti btaA protein.
XX
KW Betaine lipid production; betaine lipid compound; DGTS;
KW diacylglycerol-O-4'-(N,N,N-trimethyl) homoserine; agricultural industry;
KW phosphate-containing fertilizer; transgenic; btaA; enzyme.
XX
OS Sinorhizobium meliloti.
XX
PN WO200283844-A2.

```

XX PD 24-OCT-2002.
 XX XX 09-APR-2002; 2002WO-US011134.
 XX PF 13-APR-2001; 2001US-0283812P.
 XX PR 08-APR-2002; 2002US-00283812.
 XX XX (UNMS) UNIV MICHIGAN STATE.
 XX XX Benning C, Riekhof W, Klug R;
 XX PI WPI; 2003-058632/05.
 XX DR N-PSDB; ABS58084.
 XX XX New composition comprising an isolated and purified DNA molecule, useful
 PT for producing Betaine lipids, e.g. Diacylglycerol-0-4'-(N,N,N-trimethyl)
 PT homoserine (DGTS) for agricultural applications.
 XX XX Claim 11; Fig 29; 109pp; English.
 XX XX The present invention relates to compositions and methods for producing
 CC betaine lipids. The method comprises the expression of recombinant
 CC enzymes (e.g. from Rhodospirillum rubrum) in host cells such as
 CC bacteria, yeast, and plants to produce betaine lipid compounds including
 CC diacylglycerol-0-4'-(N,N,N-trimethyl) homoserine (DGTS). The methods and
 CC compositions of the invention are useful in agricultural applications,
 CC such that the amount of phosphate-containing fertilizer required for the
 CC growth of a particular plant is decreased. The polynucleotide sequences
 CC encoding the recombinant enzymes may be used to produce vectors which can
 CC be used to produce transgenic plants. The present sequence represents the
 CC R. sphaeroides betaine homologue from Sinorhizobium meliloti
 XX XX Sequence 416 AA;
 XX XX
 XX Query Match 8.4%; Score 292.5; DB 6; Length 416;
 XX Best Local Similarity 26.0%; Pred. No. 1e-20;
 XX Matches 102; Conservative 71; Mismatches 174; Indels 45; Gaps 16;
 QY 275 LYTQSWEDPDPMEVMEINPKDVTLTGSGCNALNLLVQAGQVSVDCNPAQSALLEL 334
 Db 45 VYQIWEDEPVDMEAMQIRPHRIVTIGSGCNMLTVLSAEPARIDVDLNPHEALNRL 104
 QY 335 KKVAIQOL-BFEDVWQVLF-EGVHPRIELEYKLAFFLSQTSNFWMS-----KRLWY 385
 Db 105 KLSAFRHLPSHKDVRFLAVEGTRTN-GQAYDVFLAPKLPDPATRAYWNGRDLTGRRRIGV 163
 QY 386 FQHGLYYQGGMGKLCWVLCQAVLVGLGKTVKRLANAPTMEORRLWDSNMLHFVNKGP 445
 Db 164 FGRNVYRTGLGRFISASHALRLHGINP--EDFVKARSMEQRFDDKLAFLP---E 217
 QY 446 KPLV-WLFVKFVSLVLFNKAFLWFGGVPGKQYALIKADGIPENTYARTMDGVAENSHV 504
 Db 218 RPVIRWITSKSSL-----FGLGIPQOQDELAS--LSREKSVAAVLRNLEKLTTC 266
 QY 505 R---KONYFYNNCLTKGFLRDN---CPTYLREAAFPATLKSQVVDNLVTSTNFMFEEEL--- 555
 Db 267 HFPLRDNYFAWQAFARYPRPDGELPPYLOASRYEAIRDN-AERVEVHHSATTELLAGK 325
 QY 556 KARTYTKVILMDHVDLMDVANELAECLAKQVAPGGIVWRS---ASLSPP--YAEILQ 610
 Db 326 PAASVDVRYLLDQDWMWTDQDLNDLWTEITRTADAGAVIFRTAAEASILPGRLSTLLD 385
 QY 611 KAGFDVRCIRATQGYMDRVNMTSSSFYMAR 642
 Db 386 QWTYDAETSRL--GAEDRSAYGGFHYIRKK 415
 RESULT 12
 ID ADW72736
 XX ADW72736 standard; protein; 416 AA.
 AC ADW72736;

XX DT 21-APR-2005 (first entry)
 XX DE Sinorhizobium meliloti betaine gene, protein.
 XX KW Phospholipid synthesis; transgenic plant; fertilizer; enzyme;
 XX KW betaine lipid.
 XX OS Sinorhizobium meliloti.
 XX PN WO2005009115-A2.
 XX PD 03-FEB-2005.
 XX PF 15-JUL-2004; 2004WO-US022789.
 XX PR 16-JUL-2003; 2003US-00620914.
 XX XX (UNMS) UNIV MICHIGAN STATE.
 XX XX Benning C, Riekhof W;
 XX PI WPI; 2005-112975/12.
 XX DR N-PSDB; ADW72735.
 XX XX New composition comprising a purified DNA having an oligonucleotide
 PT sequence encoding a protein, useful in producing Betaine lipid compounds
 PT e.g., Diacylglycerol-0-4'-(N,N,N-trimethyl) homoserine (DGTS).
 XX XX Disclosure; SEQ ID NO 33; 147pp; English.
 XX XX The invention relates to a new composition comprising a purified DNA
 CC having an oligonucleotide sequence appearing as ADW72747 (Chlamydomonas
 CC reinhardtii BtAI cDNA) or ADW72752 (Neurospora crassa BtAI coding region,
 CC encoding BtAI proteins which are DGTS (diacylglycerol-0-4'-(N,N,N-
 CC trimethyl)homoserine) synthetic enzymes. BtAI (and BtAA and BtAB) enzymes
 CC allow the replacement of phospholipids with non-phosphorus containing
 CC lipids in the cell membrane. A transgenic plant expressing the enzymes
 CC would have a reduced need for phosphate-containing fertilizer. Also
 CC included are an RNA transcribed from the purified DNA, antibodies
 CC produced from the protein, a vector comprising the DNA or nucleic acid, a
 CC host cell comprising the vector, transgenic plants comprising the vector,
 CC and a protein translated from the RNA or encoded by the nucleic acid. The
 CC composition is useful in producing Betaine lipid compounds e.g.,
 CC Diacylglycerol-0-4'-(N,N,N-trimethyl) homoserine (DGTS). The composition
 CC of the invention reduces the amount of phosphate fertilizer needed for
 CC the optimal growth of crop plants. The present sequence represents a
 CC betaine lipid synthetic enzyme.
 XX XX Sequence 416 AA;
 XX XX
 XX Query Match 8.4%; Score 292.5; DB 9; Length 416;
 XX Best Local Similarity 26.0%; Pred. No. 1e-20;
 XX Matches 102; Conservative 71; Mismatches 174; Indels 45; Gaps 16;
 QY 275 LYTQSWEDPDPMEVMEINPKDVTLTGSGCNALNLLVQAGQVSVDCNPAQSALLEL 334
 Db 45 VYQIWEDEPVDMEAMQIRPHRIVTIGSGCNMLTVLSAEPARIDVDLNPHEALNRL 104
 QY 335 KKVAIQOL-BFEDVWQVLF-EGVHPRIELEYKLAFFLSQTSNFWMS-----KRLWY 385
 Db 105 KLSAFRHLPSHKDVRFLAVEGTRTN-GQAYDVFLAPKLPDPATRAYWNGRDLTGRRRIGV 163
 QY 386 FQHGLYYQGGMGKLCWVLCQAVLVGLGKTVKRLANAPTMEORRLWDSNMLHFVNKGP 445
 Db 164 FGRNVYRTGLGRFISASHALRLHGINP--EDFVKARSMEQRFDDKLAFLP---E 217
 QY 446 KPLV-WLFVKFVSLVLFNKAFLWFGGVPGKQYALIKADGIPENTYARTMDGVAENSHV 504
 Db 218 RPVIRWITSKSSL-----FGLGIPQOQDELAS--LSREKSVAAVLRNLEKLTTC 266
 QY 505 R---KONYFYNNCLTKGFLRDN---CPTYLREAAFPATLKSQVVDNLVTSTNFMFEEEL--- 555
 Db 267 HFPLRDNYFAWQAFARYPRPDGELPPYLOASRYEAIRDN-AERVEVHHSATTELLAGK 325
 QY 556 KARTYTKVILMDHVDLMDVANELAECLAKQVAPGGIVWRS---ASLSPP--YAEILQ 610
 Db 326 PAASVDVRYLLDQDWMWTDQDLNDLWTEITRTADAGAVIFRTAAEASILPGRLSTLLD 385
 QY 611 KAGFDVRCIRATQGYMDRVNMTSSSFYMAR 642
 Db 386 QWTYDAETSRL--GAEDRSAYGGFHYIRKK 415

Db 267 HFPLRDNYPFAWQAFARYPRDEGELPPYLQASRYEAIRDN-AERVEVHHASFTLLAGK 325
 QY 556 KATYTKVILMDHVDLMDPVAELAECLAKQVAPGIVWRS---ASLSP--YAEILQ 610
 Db 326 PAASVDRYVLLDAQDMWTDQQLNDLWTEITRTADAGAVVIFRTAAEASILPGRSLTLLD 385
 QY 611 KAGFDVRCIRRAATQGVMDRVNMYSSFYMARCK 642
 Db 386 QWYDAETSURL--GAEDRSAYGGFHYIRKK 415

RESULT 13

ADT04703

ID ADT04703 standard; protein; 2284 AA.

XX AC ADT04703;

DT 30-DEC-2004 (first entry)

XX House mouse DNA polymerase epsilon full-length protein - SEQ ID 58.

DE DNA replication; cytostatic; antianaemic; CNS; dermatological; endocrine;

XX respiratory; urological; cancer; tumour; anaemia; leukaemia;

KW malignant lymphoma; nervous system; immune disorder; skin disorder;

KW endocrine disorder; respiratory disorder; digestive disorder;

KW urinary disorder; genital disorder; circulatory disorder; motor disorder;

KW plant disease; DNA polymerase epsilon; POLE; POL2; enzyme.

XX Mus musculus.

XX WO2004087960-A2.

XX 14-OCT-2004.

XX 26-MAR-2004; 2004WO-JP004378.

XX 28-MAR-2003; 2003JP-00092898.

PR 10-OCT-2003; 2003US-00684141.

XX (NEOM-) NEO-MORGAN LAB INC.

PA (FURU//) FURUSAWA M.

XX Furusawa M;

XX WPI; 2004-737733/72.

DR N-PSDB; ADT04702.

XX Regulating a conversion rate of a hereditary trait of a cell comprises

PT regulating an error-prone frequency of gene replication of the cell.

XX Example 3; SEQ ID NO 58; 488pp; English.

PS The invention relates to a novel method for regulating the conversion

XX rate of a hereditary trait of a cell. The method comprises regulating the

CC error-prone frequency of gene replication of the cell. The method of the

CC invention has cytostatic, antianaemic, CNS, dermatological, endocrine,

CC respiratory and urological applications. The polymerases of the invention

CC may be useful for regulating the conversion rate of a hereditary trait of

CC an organism or for producing an organism having a regulated hereditary

CC trait, where the polymerases have a different error-prone frequency. Such

CC an organism may be resistant to an environment. The method and

CC compositions may be useful for treating a disease or disorder,

CC particularly cancer or tumours, anaemia, leukaemia, malignant lymphomas,

CC diseases or disorders related to nervous, immune, skin, endocrine,

CC respiratory, digestive, urinary, genital, circulatory and motor systems,

CC or diseases and disorders of plants. The current sequence is that of a

CC DNA polymerase epsilon (POLE;POL2) full-length protein of the invention.

XX Sequence 2284 AA;

SQ

Query Match

Best Local Similarity 3.6%; Score 123; DB 8; Length 2284;

Matches 141; Conservative 79; Mismatches 217; Indels 240; Gaps 35;

QY 107 HSLCEVAKKAKAK-GWKNVQVVEADACOFAPPEGTATLITFSYS----- 150
 Db 422 HNL-----KAAAKAKGLGDFVELDPDMCRWA-TEQPTLATYSVDAVATYLYMYKVHP 476
 QY 151-----LTMIPFFH-NVI-----DQACSYLSODGLVGVA 177
 Db 477 FIFALCTIIPMEDEVLRKSGTLCALLMWQAFHANIIFPNKQEQEFNKLTDGHVLD 536
 QY 178 DFYVSG-----KYDLP--LRQMPWSRRPFWRSIFDIDNIDIGPERRAYLEQ--KL 223
 Db 537 ETVYGGHVEALSSGVFRSDIPCRFRMNPAAFDLQORVEKTMRAHEEBEKEVPVEQATNF 596
 QY 224 ERVWEQ--NTQSGIPVVPWMLRABPYVYVWIGRLPSVGHALHEERVERP-----PMPPT 273
 Db 597 QEVCEQIKTSLKDV-----NRIECLPIYHLHDVGAMYFNI 634
 QY 274 FLYTQSWEDPEPDMVMEINPKDVTLTLSGGCN-----ALNLLVQAGQVSVDCN 325
 Db 635 ILTNR-----LQPSAIVDEATCAACDNKPGASCQCKMAWQWRGEFM----- 676
 QY 326 PAQSALLELKVAIQOLEEEDVWOLFEGVHPRIEL-----YEK-KLAPP----- 370
 Db 677 PASRS--EYHRIQ-HOLESEKFPPLPEGPAPAFHLSREEQAKYKRRRLADYCRKAYKK 733
 QY 371-----LSQTSNFWSKRLWYFQHGLYQCGMGKLCWVLCQCLAVVLGLGKTVKR 418
 Db 734 IHVTKVEERLTTCORENSFYVDTVAFRRDRVEFKGLHKVKKKLSAAVEVDGASEVKR 793
 QY 419 LANAPTMESQRLWDSNMLIH-----FVYNGPKPLVWLFFVFPVSLVLFNKAVLWF 468
 Db 794 CKN-----MEILYDSLQLAHKCILNSFYGVNMRKGAR--WYSMEMAGIVCFT----- 838
 QY 469 GGGVPGKQYALIKADGIP-----ENYIARTMDGVAENSHVRKQNVFY 512
 Db 839 GANIITQARELIEQIGRPLELDTDGIWCVLPLNSFPENFVIKT-----TNAKKPLTISYP 893
 QY 513 NCLTGKFLRDNCPY-----LREAAFATLKSVDNLTSTNFFMEELKARTYTKVILMDH 568
 Db 894 GAMLNIMVKEGFTNHQYQELTEPSSITY-----VTHSENSIFFEVDG-PYLAMIL--- 942
 QY 569 VDWLDMPVANELAECLAKQVAPGGIVWRSASISPPYAEILIQKAGFDVRCIRRAATQGYMD 628
 Db 943-----PASKEGKGLKRYA---VFNEGDSL-----AEL---KGPEVK-----RRGELQ 980
 QY 629 RVNMY-SSFYMARCKGA 644
 Db 981 LIKIFQSSVFEAFIRGS 997

RESULT 14

ADCS7627

ID ADC97627 standard; protein; 212 AA.

XX AC ADC97627;

DT 01-JAN-2004 (first entry)

DE E. faecium protein sequence SEQ ID 7254.

XX Vaccine; urinary tract infection; bacteraemia; endocarditis; wound;

XX abdominal-pelvic infection.

XX Enterococcus faecium.

XX US6583275-B1.

XX 24-JUN-2003.

XX 30-JUN-1998; 98US-00107532.

XX 02-JUL-1997; 97US-0051571P.

PR 14-MAY-1998; 98US-0085598P.

```
XX      (UNWS ) UNIV MICHIGAN STATE.  
PA  
XX Benning C, Riekhof W, Klug R;  
PI  
XX WPI; 2003-058632/05.  
DR N-PADB; ABS98085.  
DR DR  
XX New composition comprising an isolated and purified DNA molecule, useful  
PT for producing Betaine lipids, e.g. Diacylglyceryl-0'-4'-(N,N,N-trimethyl)  
PT PT homoserine (DGTS) for agricultural applications.  
XX Claim 11; Fig 31; 109pp; English.  
XX The present invention relates to compositions and methods for producing  
CC betaine lipids. The method comprises the expression of recombinant  
CC enzymes (e.g. from Rhodobacter sphaeroides) in host cells such as  
CC bacteria, yeast, and plants to produce betaine lipid compounds including  
CC diacylglyceryl-0'-4'-(N,N,N-trimethyl) homoserine (DGTS). The methods and  
CC compositions of the invention are useful in agricultural applications,  
CC such that the amount of phosphate-containing fertilizer required for the  
CC growth of a particular plant is decreased. The polynucleotide sequences  
CC encoding the recombinant enzymes may be used to produce vectors which can  
CC be used to produce transgenic plants. The present sequence represents the  
CC R. sphaeroides btas homologue from Sinorhizobium meliloti  
XX Sequence 221 AA;  
SQ
```

```
Query Match          3.5%; Score 120; DB 6; Length 221;  
Best Local Similarity 24.8%; Pred.No. 0.0034;  
Matches 39; Conservative 34; Mismatches 67; Indels 22; Gaps 6;  
  
QY    75 VDLGGGTGGENVMADYYIDLAKFSIYVVDLCSHLCSEVAKKKAGKNVVVEADACQ 134  
       ::::| | | : | : | : | : | : | : | : | : | : | :  
Db     53 LEVGCGTGRNLAVIGDLYPGAR---LFGLDISAEMLATATAKLRRQRNPDAVLRVADATN 109  
  
QY   135 FA-----PPGTATLTFTSLSLTMPFPHNVIDOCASYLSODGLGVGVADFVSYGKYDIPLR 190  
       ||: | |: | |: | |: | |: | |: | |: | |: | |: |  
Db    110 FTAASFDQGDFPRIV-ISYALSVPPEWEKAVIDAAIALKPSSLHTIADFQQEGGWAPAGR 168  
               | | | | | | | | | | | | | | | | | |  
QY   191 Q--MPSRRFF---WRSIFDN-----IDI GPERRAY 218  
           | | | | | | | | | | | | | | | | | |  
Db    169 RFLOALWRPFHPVTRETLDVMEKRRAERNGAALVELSRGY 210
```

```
Search completed: March 14, 2006, 01:55:26  
Job time : 107 secs
```

[illegible]

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 14, 2006, 01:46:11 ; Search time 32.3 Seconds
(without alignments)
5037.237 Million cell updates/sec

Title: US-10-620-914-44

Perfect score: 3619

Sequence: 1 atggggctgggtcgtgcagg.....gcgcgaagaaggacaactaa 1947

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 3735138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlh
-Q=/abs/ABSSWEB spool/US10620914/runat_13032006_101936_5626/app query.fasta_1
-DB=Published Applications AA Main -OPMT=fastan -SUFFIX=n2p.rapbm
-MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1
-MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DLOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext
-HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abs05h
-USER=US10620914 @CGN 1 1 134 @runat_13032006_101936_5626 -NCPU=6 -ICPU=3
-NO MMAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA Main:

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	3463	95.7	648	US-10-620-914-45	Sequence 45, Appl
2	891	24.6	752	US-10-741-849-7021	Sequence 7021, Ap
3	850.5	23.5	908	US-10-620-914-50	Sequence 50, Appl
c 4	476.5	13.2	19662	US-10-084-846A-6	Sequence 6, Appl
5	459	12.7	19723	US-10-084-846A-5	Sequence 5, Appl
c 6	457.5	12.6	19723	US-10-084-846A-7	Sequence 5, Appl
c 7	456.5	12.6	19652	US-10-084-846A-7	Sequence 7, Appl
8	446	12.3	19608	US-10-084-846A-3	Sequence 8, Appl
9	443.5	12.3	19695	US-10-084-846A-3	Sequence 3, Appl
c 10	436	12.0	19608	US-10-084-846A-8	Sequence 8, Appl
c 11	424	11.7	19725	US-10-084-846A-4	Sequence 4, Appl

12	407	11.2	19725	4	US-10-084-846A-4	Sequence 4, Appl
c 13	383.5	10.6	19695	4	US-10-084-846A-3	Sequence 3, Appl
14	379.5	10.5	416	4	US-10-118-495-3	Sequence 3, Appl
15	379.5	10.5	416	4	US-10-620-914-3	Sequence 3, Appl
c 16	370	10.2	900	5	US-10-450-763-33892	Sequence 33892, A
17	366.5	10.1	1084	5	US-10-794-514A-336	Sequence 336, App
c 18	357	9.9	1084	5	US-10-794-514A-336	Sequence 336, App
c 19	352	9.7	599	5	US-10-450-763-33889	Sequence 33889, A
20	342.5	9.5	19662	4	US-10-084-846A-6	Sequence 6, Appl
21	341.5	9.4	437	4	US-10-437-963-132079	Sequence 132079,
22	335.5	9.3	465	4	US-10-437-963-174113	Sequence 174113,
23	336.5	9.0	416	4	US-10-118-495-29	Sequence 29, Appl
24	326.5	9.0	416	4	US-10-620-914-29	Sequence 29, Appl
25	321	8.9	415	4	US-10-118-495-41	Sequence 41, Appl
26	321	8.9	415	4	US-10-620-914-41	Sequence 41, Appl
27	320.5	8.9	19652	4	US-10-084-846A-7	Sequence 7, Appl
28	308	8.5	455	4	US-10-437-963-161457	Sequence 161457,
c 29	307	8.5	497	4	US-10-437-963-125004	Sequence 125004,
30	304	8.4	599	5	US-10-450-763-33889	Sequence 33889, A
31	303.5	8.4	4440	4	US-10-052-586-525	Sequence 525, App
32	303.5	8.4	4440	4	US-10-174-590-525	Sequence 525, App
33	303.5	8.4	4440	4	US-10-176-758-525	Sequence 525, App
34	303.5	8.4	4440	4	US-10-175-737-525	Sequence 525, App
35	303.5	8.4	4440	4	US-10-174-581-525	Sequence 525, App
36	303.5	8.4	4440	4	US-10-176-483-525	Sequence 525, App
37	303.5	8.4	4440	4	US-10-176-749-525	Sequence 525, App
38	303.5	8.4	4440	4	US-10-176-914-525	Sequence 525, App
39	303.5	8.4	4440	4	US-10-176-915-525	Sequence 525, App
40	303.5	8.4	4440	4	US-10-173-706-525	Sequence 525, App
41	303.5	8.4	4440	4	US-10-175-738-525	Sequence 525, App
42	303.5	8.4	4440	4	US-10-175-752-525	Sequence 525, App
43	303.5	8.4	4440	4	US-10-176-482-525	Sequence 525, App
44	303.5	8.4	4440	4	US-10-176-757-525	Sequence 525, App
45	303.5	8.4	4440	4	US-10-176-913-525	Sequence 525, App

ALIGNMENTS

RESULT 1

US-10-620-914-45
; Sequence 45, Application US/10620914
; Publication No. US20040093639A1
; GENERAL INFORMATION:
; APPLICANT: Benning, Christoph
; APPLICANT: Riekhof, Wayne
; TITLE OF INVENTION: Compositions and Methods for the Production of Betaine Lipids
; FILE REFERENCE: MSU-07789
; CURRENT APPLICATION NUMBER: US/10/620,914
; CURRENT FILING DATE: 2003-07-16
; PRIOR APPLICATION NUMBER: 10/118,495
; PRIOR FILING DATE: 2002-04-08
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 45
; TYPE: PRT
; ORGANISM: Chlamydomonas reinhardtii
US-10-620-914-45

Alignment Scores:
Pred. No.: 1.73e-204 Length: 648
Score: 3463.00 Matches: 648
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 95.7% Indels: 0
DB: 4 Gaps: 0

US-10-620-914-44 (1-1947) x US-10-620-914-45 (1-648)

Qy 1 ATGGGTCGGGTCTGTCAGCGCGCGCTGCAGGTACACCAAGAGACTTCTCCCTCGAG 60
|||||
Db 1 MetGlySerGlyArgAspGlyArgProAlaSerTyrThrIlyAsnPhenSerLeuGlu 20

QY	61	AAGCTCAAGCTCAGCAGCATGAAGAGTACCTGACCGTTCTTGCGCCATATGTGGTTCGGC	120
DB	21	LysLeuLysLeuSerSerMetLysAspAspLeuThrValIleArgHisMetIrpPheGly	40
QY	121	AGCAAGAGGGCGATGATCATCGTGTCTCGCTCGAGAGCTTCTACGGGCCCCAGGCCGT	180
DB	41	SerLysLysGlyAspAspHisAlaAlaArgLeuGluSerPheIrpGlyProGlnAlaAla	60
QY	181	GCCTTTCCTGCGCCCTGGCGGAGCGCTCGAACTCATCTGGGTTCACCTGGGTGGTGGC	240
DB	61	AlaPheAlaAlaArgLeuAlaGluArgSerAsnLeuIleIrpValAspLeuGlyGlyGly	80
QY	241	ACTGGGAGAGTGTGCATATGATGGCTGATTACATCGACTGGCGAAGTTCGAAGTCCATC	300
DB	81	ThrGlyGluAsnValAspMetAlaAspTyrIleAspLeuAlaLysPheLysSerIle	100
QY	301	TACGTGTGCACCTGTGCCACTCGCTGTGCGAGTGGCCAAAGAAGCGAAGCGCAAG	360
DB	101	TyrValValAspLeuCysHisSerLeuCysGluValAlaLysLysLysAlaLysAlaLys	120
QY	361	GGCTGGAAGATGTCAGGTCGTGGAGGCGAGCGCTTGCCAAATTTGGGCCCCCTGAGGGC	420
DB	121	GlyTrpLysAsnValGlnValValGluAlaAspAlaCysGlnPheAlaProProGluGly	140
QY	421	ACCGCGAGCTCATCACCTTCTCTACTCGCTCACGATGATTCACCGTTCACCAACGTC	480
DB	141	ThrAlaThrLeuIleThrPheSerTyrSerLeuThrMetIleProProPheHisAsnVal	160
QY	481	ATCGACGAGGTTGCTCGTACCTCTCTGCTCCAAAGACGGCTGTGGCGCTGCCGACTTCTAC	540
DB	161	IleAspGlnAlaCysSerTyrLeuSerGlnAspGlyLeuValGlyValAlaAspPheTyr	180
QY	541	GTGAGCGGCAAGTAGACAGCTGCCCTCGCGCAGATGCCCTGTGTGGCGCGTTCCTCTCG	600
DB	181	ValSerGlyLysTyrAspLeuProLeuArgGlnMetProTrpSerArgPhePheTrp	200
QY	601	CGATCGATCTTCGACATCGACATTTGATCGGCCCGAGCGCGCGCTACCTCGGAG	660
DB	201	ArgSerIlePheAspIleAspAsnIleAspIleGlyProGluArgArgAlaTyrLeuGlu	220
QY	661	CAGAACTGGAGCGCGTGTGGAGCAAAACACCGAGGTTTCGATCCCTACGTGCCGTGG	720
DB	221	GlnLysLeuGluArgValTrpGluIleAsnThrGlnGlySerIleProTyrValProTrp	240
QY	721	CTGCGCGCCCTACTACTGTGTGATTGGCGCTGCGCCAGCGTGGCCACGCCCTGCAC	780
DB	241	LeuArgAlaProTyrTyrValTrpIleGlyArgLeuProSerValGlyHisAlaLeuHis	260
QY	781	GAGGAGCGGTGGAGCGCGCCCATCTTCCCGCCACCTTCTGTGTACACGCAAGTGGTG	840
DB	261	GluGluArgValGluArgProPheProPheProThrPheLeuTyrThrGlnSerIrp	280
QY	841	GAGGACCCCGAGCGCGATATGGAGGTGATGGAGATCAACCCCAAGGACACGGTGTGACC	900
DB	281	GluAspProGluProAspMetGluValMetGluIleAsnProLysAspThrValLeuThr	300
QY	901	CTGACTAGCGCGGTGCAATGCCCTGAACTGTGTGTGTGAGGGGCGGCCACAGTGGTG	960
DB	301	LeuThrSerGlyGlyCysAsnAlaLeuAsnLeuLeuValGlnGlyAlaGlyGlnValVal	320
QY	961	TCGGTGGACTGCAACCCCGCCAGTGGCGCTTCTGGAGCTGAGAGGTGGCCATTTCAG	1020
DB	321	SerValAspCysAsnProAlaGlnSerAlaLeuLeuGluLeuLysLysValAlaIleGln	340
QY	1021	CAGCTGGAGTGTGAGGACGTGGCAGCTGTTCGGCGAGGCGGTGACCCCGCATTTGAG	1080
DB	341	GlnLeuGluPheGluAspValTrpGlnLeuPheGlyGluGlyValHisProArgIleGlu	360
QY	1081	GAGCTGTACGAGAAGACTCGCGCCCTTCTGTGCGAAACCGACCCACAACCTTGTGTC	1140
DB	361	GluLeuTyrGluLysLysLeuAlaProPheLeuSerGlnThrSerHisAsnPheTrpSer	380

RESIST 2

```

US-10-741-849-7021
; Sequence 7021, Application US/10741849
; Publication No. US20050019931A1
; GENERAL INFORMATION:
; APPLICANT: Roemer, Terry
; APPLICANT: Jiang, Bo
; APPLICANT: Boone, Charles
; APPLICANT: Bussey, Howard
; TITLE OF INVENTION: Nucleic Acids Encodin
; TITLE OF INVENTION: Use
; FILE REFERENCE: 10182-023-999
; CURRENT APPLICATION NUMBER: US/10/741,849
; CURRENT FILING DATE: 2003-12-19
; PRIOR APPLICATION NUMBER: US 60/434,832
; PRIOR FILING DATE: 2002-12-19
; NUMBER OF SEQ IDS: 8000

```

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 7021

; LENGTH: 752

; TYPE: PRT

; ORGANISM: Candida albicans

US-10-741-849-7021

Alignment Scores:

Pred. No.: 2,69e-46 Length: 752
 Score: 891.00 Matches: 229
 Percent Similarity: 46.6% Conservative: 98
 Best Local Similarity: 32.7% Mismatches: 232
 Query Match: 24.6% Indels: 142
 DB: 5 Gaps: 21

US-10-620-914-44 (1-1947) x US-10-741-849-7021 (1-752)

```

QY 151 CTGGAGAGCTTCTAGCGGCCCGCCGCTGCTTT----- 186
Db 70 LeuGluSerPheTyrLysAsnGlnAlaHisIleTyrAspAsnThrArgGluPheLeuLeu 89
QY 187 -----GTCGCCCGCTTGGCGCGAGCGCTCGAACCTC 216
Db 90 LysGlyArgGlnGluCysLeuArgLeuAlaIleSerHisLeuProLysLysLeu 109
QY 217 ATCTGGGTGACCTGGGTGGGCTGGGAGAAATGTCGATATGATGGCTGATTACATC 276
Db 110 IleTrpIleAspIleGlyGlyThrGlySerAsnIleGluPheMetAspGluIleSer 129
QY 277 GACCTGGCGAG---TTCAAGTCCATCTACGTGTGGTGGACCTGTCCTGTCGGAG 333
Db 130 LysIleSerGluAsnPheLysAlaValTyrLeuValAspLeuSerProSerLeuCysGlu 149
QY 334 GTGGCCAAAGAAAGCGCAAGCGCTGGAAGATGTCCAGGTGCTGGAGCGCCGAC 393
Db 150 ValAlaLysAlaArgPheGluAlaHisIleGluTrpThrAsnValHisValLeuValAlaAsp 169
QY 394 GCTTGCCAAATTTGGCCCCCTGAGGGCACCGGACGCTCATCCTCTCTCTCTCTCTC 453
Db 170 AlaCysAspPheThrIleAspTyrAspSerAlaAspLeuIleThrPheSerTyrSerLeu 189
QY 454 ACGATGATTCACCGTCCCAACGTCATCACCAGCGTCTGCTGCTACCTGTCCTCCAGAC 513
Db 190 SerMetIleProThrPheAsnAlaAlaIleAspAsnAlaValSerLysLeuAspMetGlu 209
QY 514 GGCCTGGTGGCGTGGCGACTTCTACGTGAGC-----GGCAAGTAC 555
Db 210 GlyIleIleAlaThrValAspPheGlyIleGlnSerSerAspThrSerMetGlyArgIle 229
QY 556 GAC-----CTGCCCTCGCCAGATGCCGTGGTGGCGCGCTTCTCTGGCGA 603
Db 230 AsnThrValGlyLeuValAsnArgAspIleProTrpIleLeuArgAsnPheTrpArg 249
QY 604 TCGATCTTCGACATCGACAATGACATCGGCCCGCGCGCGCTACCTACCTGGAGCAG 663
Db 250 IleTrpPheGluAlaAspLysValPheLeuAspSerSerArgArgAsnTyrLeuGluTyr 269
QY 664 AAGCTGGAGCGGTGGGAGCAGACACACCAG-----GGTTCGATCCCTTAC 711
Db 270 LysPheGlyThrValLysSerLeuAsnSerTyrAsnLysAlaLeuGlyLysIle----- 287
QY 712 GTGCCGTGGTGGCGCCCTTACTAGTGTGGATGGCGCCCTCGCCAGCGTTGGCCAC 771
Db 288 -----ProTyrTyrIleTrpIleGlyCysAspLysSerLysSerHis 301
QY 772 GCCTTGACGAGGAG-----GGCGTGGAGCGCGCCCGCATGTTCCCGCCCGCACC 819
Db 302 ThrIleLeuGluArgLeuAsnCysLeuAlaThrGluSerProTyrLeuAlaProThrThr 321
QY 819 ----- 819
Db 322 ThrProIleAlaAsnGlnLeuGluAspIleProIleSerLysGlyHisGluAlaLeu 341

```

```

QY 819 ----- 819
Db 342 IleAsnLeuGlnLysAsnLeuProTyrProSerMetTyrTyrGlnLysGluTyrTrpArg 361
QY 820 -----TTCCTGTACACG-831
Db 362 ValTyrTyrAspGluMetAsnProLeuTyrGluGlnPheLysAsnGlnTyrIleTyrAla 381
QY 832 CAGTCGTGGGAGGACCCCGAGCGGATATGAGGTGATGGAGATCAACCCCAAGACACG 891
Db 382 PheThrTrpGluAspProArgGluAspHisLysLeuLeuAsnPheThrSerAspAspThr 401
QY 892 GTGCTGACCCCTGACTAGCGCGGCTGCATCCCTGAACTG-----CTGGTGACGGG 945
Db 402 ValLeuAlaIleThrSerAlaGlyAspAsnIleLeuSerTyrAlaSerLeuProThrPro 421
QY 946 GCGCGCAGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1005
Db 422 ProLysLysIleHisAlaValAspLeuAsnProCysGlnAsnHisLeuLeuGluLeuLys 441
QY 1006 AAGTGGCCCATTCACGAGCTGGAGCTTTGAGACGCTGGGACGCTGTTCCGGGAGGCGGTG 1065
Db 442 LeuAlaSerPheArgCysLeuSerGlnGluGlnIleTrpSerMetPheGlyGluGlyLys 461
QY 1066 CACCGCGCATTTGAGGAGCTGTACGAGAAGCTGGCGCCCTTCCTGTCGCAACACGAC 1125
Db 462 IleGluAsnPheAsnAspLeuLeuIleAspThrLeuAlaProHisMetSerSerAsnAla 481
QY 1126 CACAACCTCTCTGGTCCAGCGC-----CTCTGGTACTTCCAGCAGCGCTCTACTACCAG 1179
Db 482 PheGlnTyrTrpMetAspLysGlyProLysThrPheSerGlyLysGlyLeuTyr----- 499
QY 1180 GCGCGCATGGCGAAGCTGTGCTGGTGTCTGAGTGC-----CTGGCGCTGTGTGCTG 1230
Db 500 ---AspThrGlyPheSerArgTrpAlaLeuArgLeuSerArgTyrValPheLysValCys 518
QY 1231 GGACTGGGCAAGACCGCTCAAGCGCTCGCCNACCGCCCAACATGAGGAGGAGCGCGCT 1290
Db 519 GlyValSerLysTyrValGluGluCysAlaAlaThrThrMetGluGluGlnLeuArg 538
QY 1291 CTGTGGGACGACACATGCTCATCCACTCTCGTGAAGAACGGCGCCCGCTGTGTGG 1350
Db 539 IleTrpAsnGluHisLeu-----LysProThr----- 547
QY 1351 CTGTTCTCTCAAGTTCGTG---AGCCTGCTGCTTCAACAGGCGCTGCTGTGTTCGGC 1407
Db 548 LeuPheAsnProValValGlySerLeuLeuValGlyAsnProMetPheLeuTrpLysAla 567
QY 1408 GCGCGCTGCGCGCAAGCAGTACGCGCTGATCAAGCGGAGCGGATCCCCCATTTGAGAAC 1467
Db 568 LeuGlyValProAlaAsnGlnAlaLeuMet-----GlyProSerValIleLys 584
QY 1468 TACATCGCGCGCACCATGGAGCGGTGGCGGAGAACTCGCACGTCGCGACGACGACTAC 1527
Db 585 TyrValValAspThrLeuAspProIleLysArgSerMetIleSerAsnAspAsnTyr 604
QY 1528 TTCTACTACAACTGCTCAGCGCAAGTCTCTGCGCACAACTGCCCCCAGCTGCTGCGC 1587
Db 605 PheTyrTyrLeuCysMetMetGlyArgTyrThrLysAsnAsnCysProAspTyrLeuThr 624
QY 1588 GAGCGCGCTTGGCCACCCCTCAAG-----AGTGGCGTGGTG 1623
Db 625 ThrLysGlyPheAsnArgLeuSerSerThrAlaAlaThrAlaSerGlySerSerProIle 644
QY 1624 GACNACTGACCGCTCTCCACCACTTCTTCATGGAG-----GAGCTCAAGCGCGC 1674
Db 645 AspAsnLeuArgIleHisThrAspThrLeuAsnGluValPheGlyArgLeuLysGluLys 664
QY 1675 ACCTACACCAAGGTGATTCGTGAGCACGCTGACCTGGCTGGAT-----ATGCC 1725
Db 665 SerIleThrIleAlaIleIleMetAspHisMetAspTrpPheAspProAsnGlyArgAsp 684
QY 1726 GTGGCCAAACGAGCTGGCGGAGTGCCTGCGCCCAAGCAGGTGCGCGCGGCGCATCGTCATC 1785

```



```

QY 1351 CTGTCGTCACAGTTCGTGAGC---CTGGTGTCTTCAACAAGCGCGTGTGTGGTTCGGC 1407
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 667 LeuLeuAsnArgPheValAsnGlyLeuValLeuSerSerAspAlaPheLeuTrpSerAla 686
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1408 GCGCGCGTGGCGGCAAGCAGTACGCGTGTATCAAGCGCGAC-----1449
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 687 LeuGlyValProLysAsnGlnValAlaMetIleGluAlaAspTyrHisArgArgSerIle 706
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1450 -----GGCATCCCATTCAGAAC 1467
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 707 SerSerSerThrThrProSerSerLysGluLysProSerArgAlaGluAlaIleLeuHis 726
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1468 TACATCGCGGCACCATGACGGCGTGGCGGAGAACTGCGAGTGGCGCAAGCAAGACTAC 1527
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 727 TyrThrThrSerThrLeuAspProValLeuSerThrSerHisLeuAlaSerAspAsnPro 746
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1528 TTCTACTACAACTGCTCAGCGCAAGTTCCTGCGCGCAACTGCGCCACCTACTCTGCGC 1587
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 747 TyrTyrLeuValCysValLeuGlyGlnTyrThrArgGlnCysHisProAspTyrLeuSer 766
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1588 GAGCGCGCTTCGCCACCTCAAGAGT---GCGGTGTGGACAACCTGACCGTCTCCACC 1644
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 767 ProAlaAlaHisSerIleLeuSerAlaProGlyAlaPheAspGlyLeuArgIleHisThr 786
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1645 AACTTCTTCATGAGGAGCTC-----AAAGCGCGACCTACACCAAGGTGATTCTG 1695
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 787 AspGluIleGlnGluValLeuAlaArgPheGlnProGlyThrLeuThrValAlaVal 806
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1696 ATGACACACGCGACTGCTGGATATGCC-----1725
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 807 MetAspSerMetAspTyrPheAspProSerProGluGluGluLysGluGlyArgGly 826
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1726 GTGGCCAAAGAGTGGCGCGAGTGCCTGCCCAAGCAGGTGGCGCGCGCATCGTCATC 1785
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 827 LysAlaArgGlnValArgArgLeuAsnArgAlaLeuLysValGlyGlyValLeu 846
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1786 TGGCGCTCGGCTCCTCAGCGCGCTACCGCGAGTATCCAGAACGGCGGTTCGAC 1845
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 847 LeuArgSerAlaGlyValGluProTrpTyrValArgValPheValGluGluGlyPheGly 866
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1846 GTGCGCTGCATC-----CGCGCGCCACTCAGGGCTACATGACGCGGTC 1890
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 867 AlaArgValGlyCysArgGluSerGlyArgGlyAspGlnGluCysIleAspArgVal 886
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1891 AACATGTACAGCTCTCTACATGCGCGCGCGG 1923
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 887 AsnMetTyrAlaSerCysTrpIleLeuGluLys 897
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 4

```

US-10-084-846A-6
; Sequence 6, Application US/10084846A
; Publication No. US20040006026A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUHLNBERG, AGNES
; APPLICANT: TREPFER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; CURRENT APPLICATION NUMBER: US/10/084, 846A
; PRIOR FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: Patent In Ver. 3.2
; SEQ ID NO 6
; LENGTH: 19662
; TYPE: PRT
; ORGANISM: Streptomyces viridochromogenes
; FEATURE:

```

```

; OTHER INFORMATION: Protein 1: amino acid sequence encoded by coding strand 2
; OTHER INFORMATION: Start codon: gga, Start position: nucleotide 1.
US-10-084-846A-6

```

Alignment Scores:

```

Pred. No.: 1,63e-20 Length: 19662
Score: 476.50 Matches: 262
Percent Similarity: 31.6% Conservative: 42
Best Local Similarity: 27.3% Mismatches: 272
Query Match: 13.2% Indels: 385
DB: 4 Gaps: 56

```

US-10-620-914-44 (1-1947) x US-10-084-846A-6 (1-19662)

```

QY 1945 AGTTGTCTCTTCTGCGCCCTCCGCGCGGCGCATGTAGAGGAGCTGTACATGTTGACGC 1886
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 12392 ThrCysAlaGlyArgArgProGlyArgAlaProAlaArgArgAlaCys-----Arg 12408
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1885 GGTCTCATGTAGCCTGAGTGGCGCGGATGCGAGCGCACGTGGAAGC-----CCGCT 1832
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 12409 GlyGlySerProProArgSerArg---ArgCysArgProArgArgSerCysArgProGly 12427
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1831 TCTGATCAGCTCGCGCTAGCGCGGCTGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGG 1772
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 12428 ThrArgSerArgArgArgGlySerHisGlyThrArgAlaArgArg-----Arg 12444
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1771 CCGCGCAACCTGCTGGCCAGGC-----ACTCGCCAGCTCGTGGCCACGCCCATAT 1718
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 12445 ProAlaArgAlaGlyArgProAlaGlyCysArgArgProGlyArgThrProArgArgLeu 12464
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1717 CCAGCCAGTCCAGTGTTCATCAGAAATCATCTGTTGTAGTGGCGCGCTTTGAGCTCCT 1658
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 12465 ProArgLysProGlySerProProArgSerPro-----12475
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1657 CCATGAAGAAGTGGTGGAGACGCTCAGGTGTCCACCCACCCACTCTTGGAGGTGG--- 1601
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 12476 -----ArgArgGlnSerArgProProIlePheArgProProTrpAsn 12489
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1600 -----CGAAGC-----CCGCTCGCGCAGGTAGTGTCCACCCACCCACTCTTGGAGGTGG--- 1565
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 12490 ProValHisArgArgGluGlnPheTyrTrpProGlyProValValTrp-----12506
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1564 CGCGCAGAACTTCCCGTGAGGCGAGTGTAGTAGAAGTAGTCTGTGTGCGCAGCTCGC 1505
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 12507 -----ProArgTyrAlaSerAlaAlaAla 12514
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1504 AGTTCTCGCGCACCGCTCCA-----TGGTGGCGCGCATGT-----1469
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 12515 SerSerGlnProArgProProIleAlaProSerTyrProThrProCysGlyGlnArg 12534
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1468 -----AGTTCTCAATGG-----1457
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 12535 GlnIleArgSerPheGlnGlyAsnSerArgTrpHisArgMetSerProProAsnAsnAsp 12554
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1456 -----GGATCCCGT-----1448
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 12555 ArgThrGlyValPhePheArgPheSerLeuGlyPheArgAlaGlyTyrArgGlyProPhe 12574
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1447 ---CCGCTTGTATCAGCGCTAGCTGTGCGCGCAGCGCGCGCGCGCGCGCGCGCGCG 1391
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 12575 ThrProProSerProAlaAspMetSerTyrGlyValSerAlaPheArgThrHisGlyArg 12594
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1390 CCTTGTGTAAGAGACCCAGC-----1370
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 12595 ProProSerGlyValProGlyValArgSerGlyAspArgArgCysArgTrpTrpAlaAla 12614
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1369 TCACGAACTTGAGCAACAGCACCGCGCTGGCGCGCTTCTTCCAGAGTGGATGA 1310
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 12615 SerArgArgCysTrpProAlaArgProGlyArgGlyAlaSerProAlaGlyGlyGlyPro 12634
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1309 GCATGTGTGTCCTCCACAGCGCGCTGCTCTCCATTG-----1271
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 12635 ProProSerSerProSerArgAlaProAlaGlyGlyGlyGluArgAlaGlyPro 12654
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

QY 1270 TGGCGCGCTTGGCAGGCGCTTACGGTCTTCCCGCTCCAGCACACCGCCAGGCACT 1211
DB 12655 TrpGlyArgValAlaSerProAlaArgSerAlaArgAlaProAlaArgAlaProGlyAsn 12674
QY 1210 GCAGCA-----CCC 1202
DB 12675 AlaAlaProProAlaThrAspProArgThrProArgProProArgProArgProArgPro 12694
QY 1201 AGCAGAGTTGCCATGCCGCG-----CCTGGTAGTACAGGCGCGCTGCTGGA 1157
DB 12695 AlaArgArgCysProAlaArgProAlaAlaArgProProGlySerAlaGlyArgAlaGly 12714
QY 1156 AGTACCAGAGCGCT----- 1142
DB 12715 SerProGlyValProProArgProProGlyArgAlaArgCysArgArgGlyAsp 12734
QY 1141 -----TGG 1139
DB 12735 ArgAlaArgProArgProSerProProValAlaCysArgArgArgSerArgArgArg 12754
QY 1138 ACCAGAGTTGGCTGGTTT-----GCCAGAGGAAG 1106
DB 12755 AlaArgProCysGlyAlaAlaGlyProArgArgProAlaArgSerThrSerArg 12774
QY 1105 GCGCAGCTCTCTCT---CGTACA-----GCTCTCAATGGCGGCTCA 1064
DB 12775 ArgProAlaAlaGlyProArgThrArgArgAlaArgAlaProGluArgAlaValHis 12794
QY 1063 CGCCCTCCGCGAAGCTGCC-----ACAGTCTCTCAA 1031
DB 12795 ArgAspArgArgThrValAlaArgSerAsnArgAsnArgSerArgValArgProPro 12814
QY 1030 ACTCAGCTGTGTAATGGCGCTCTTTCAGCTCCAGAAAGCGCCGACTCGCGG---CGT 974
DB 12815 ArgProAlaAlaGlyArgProGlyArgGlyArgProValArgAlaGlyAlaArgProGly 12834
QY 973 TGCAGTCCACGACACACCTGGCGG-----CCCTTCGCCAGCAGAGTTTCAGGGCAT 920
DB 12835 ProAlaAlaThrAlaProProAspArgArgGlyProAlaAlaArgProSerGlyProArgHis 12854
QY 919 ---TGCAGCGCGCTAGTACAGG---TCAGCACGCTGTCTCTGGGTGATCTCCATCA 866
DB 12855 GlyArgAlaAlaArgGlySerGlyArgArgArgProArgProSerGlyArgProAlaAla 12874
QY 865 CTCCATATCCGGCTCGGGCTCTCCACGACTGCTGTGTACAGGAAGTGGCGGGAACA 806
DB 12875 ProProArgPro---ArgGlyArgPro-----GlySer 12884
QY 805 TGGCGGCGCTCCAGCGCTCTCTGTGCGAGGCGGTGGCCACGCTGGCGAGCGGCCAA 746
DB 12885 AlaAlaAspProProSerAlaPro-----GlyGlySer 12895
QY 745 TCCACAGGTAGTAGGGGGCGCGCAGCCAGCGTAGGGGATCGAACCTCGGTGTTCT 686
DB 12896 SerSerCysProGlyGlyArgGlyAlaLeuArgProArgArgProGly---Thr 12914
QY 685 GTCCACACGCGCTCCAGCTTCTGCTCCAGTAGG---CGCGGCGCTCGGGG--- 635
DB 12915 ProProAlaArgProGlyArgSerProArgGlyArgGlyCysProGlyThrGly 12934
QY 634 -----CGATGTCATGTTGT-----CGATGT-----CGAAGA 608
DB 12935 GlyCysArgArgProAlaCysCysAlaThrAlaAlaGlyArgCysProArgArgArg 12954
QY 607 TCGATC-----GCCAGAGAAACGGCGCACACAGG----- 578
DB 12955 AlaCysProAlaAlaCysGlySerArgArgArgGlyArgAlaArgProLeuAlaProArg 12974
QY 577 -----GCATCTGGCGGAGGCGGCGGCTGCTACTTGC----- 548
DB 12975 ArgTrpAsnAlaThrAlaGlyGlyAlaAlaAlaArgGlyArgAlaAlaGlyLeuArgGlySer 12994

QY 547 -----CGTCACTAGTAGCTCGGCAACGCCACCCAGGC 515
DB 12995 ArgSerAlaGlyAlaProThrSerProArgSerArgGlySerGlyArgProArg 13014
QY 514 CGTCTTGGGACAGTACGAGCAACCTGGTGCATGACGTTGT-----GGAACGGTGAA 461
DB 13015 ArgArgGlyArgSerGlyProArgAlaGly-----CysAlaValAlaSerValArg 13031
QY 460 TCATCGTAGCGAGTAGAGAGAGGTGATGAGCGTCGCGG-----TGCCCTCAG 413
DB 13032 ArgThrArgProProArgArgValArgProAlaSerArgGlySerSerAlaArgProArg 13051
QY 412 GGGCGCAAAATGGCAAGCGTCGCGCT-----CCACGACCTGGACATCTCT----- 368
DB 13052 GlyAlaGlyThrAlaArgArgThrProGlyArgProArgProArgSerArgArg 13071
QY 367 -----TCCAGCCCTTGG 356
DB 13072 ArgAlaAlaProGlyCysAlaGlySerProProSerThrArgArgArgSerProTrp 13091
QY 355 CCT-----TCGCTCTCT-----TCTTGGCCA----- 335
DB 13092 ProAlaGlyProArgAlaProSerCysArgAlaAlaCysAlaValArgTrpAspSerAla 13111
QY 334 -----CCTCGCACAGCG-----AGTGGCACAGGTCGACCA 305
DB 13112 ArgArgArgProArgAlaAlaGlyArgAlaGlyGlyProCysAlaGlyCysGlyArgPro 13131
QY 304 CGT----- 302
DB 13132 ArgArgArgProAlaValArgProAlaProArgArgSerSerGlyArgArgProSer 13151
QY 301 -----AGATGGACTTGAACCTCGCCAGGTCGATGTAAT 269
DB 13152 ArgProArgArgGlyHisArgArgSerArgTrpThr---SerSerArgThrArgTrpPro 13170
QY 268 CAGCATCATATCGACATTTCTCCCGAGTGC----- 239
DB 13171 ArgProPro---ArgHisCysProArgAlaAlaAlaArgProArgArgAlaAspPro 13189
QY 238 -----CACCCAGGTCACCCAGAGTAGAGTTTCAGC-----GCTCGGCCAGGC 194
DB 13190 SerProSerArgHisAlaProArgProArgAlaAlaAlaAlaThrTyGlyArgSerAla 13209
QY 193 GGGCAGCAAGCGCAGCGCTGGGCGCTAGAGCTCTCCAGGCGAGCAGCGTGTATCAT 134
DB 13210 GlyProArgSerArgSerAlaProArgArg-----ArgArgArgSerHis 13226
QY 133 CGCCCTTCT-----TGTCGCGAACCCACA 110
DB 13227 ArgProSerGlyArgArgArgAlaArgArgProValThrAlaThrAlaArgAla 13245
QY 109 TATGCGCGAGAACCG-----TCAGGT-----CATCTTTCATGCTGCTGAGCTTGA 65
DB 13247 ArgAlaAlaSerArgArgSerTrpSerGlyArgArgArgProArgCys----- 13262
QY 64 GCTTCTCCAGGAGAGTCTTCTTGTGTAGCTCGCAGGCGCGCTCAGCACCCGAGCC 5
DB 13263 -----GlyArgProProAlaAlaSerAlaGlyAlaArgThrArgSerProProThr 13279
QY 4 CCA 2
DB 13280 Pro 13280
RESULT 5
US-10-084-846A-5
; Sequence 5, Application US/10084846A
; Publication No. US20040006026A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUHLNBERG, AGNES
; APPLICANT: TREFFER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS

Sequence 5, Application US/10084846A
Publication No. US20040006026A1
GENERAL INFORMATION:
APPLICANT: WEITNAUER, GABRIELE
APPLICANT: MUHLENWEG, AGNES
APPLICANT: TREFFER, AXEL
APPLICANT: BECHTHOLD, ANDREAS
TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
FILE REFERENCE: 1974-005
CURRENT APPLICATION NUMBER: US/10/084,846A
CURRENT FILING DATE: 2003-02-25
PRIOR APPLICATION NUMBER: PCT/EP01/09815
PRIOR FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: DE 101 09 166.4
PRIOR FILING DATE: 2001-02-25
NUMBER OF SEQ ID NOS: 120
SOFTWARE: PatentIn Ver. 3.2
SEQ ID NO 5
LENGTH: 19723
TYPE: PRT
ORGANISM: Streptomyces viridochromogenes
FEATURE:
OTHER INFORMATION: Protein 3: amino acid sequence encoded by coding strand 1.
OTHER INFORMATION: Start codon: atc, Start position: nucleotide 3.
US-10-084-846A-5

Alignment Scores:
Pred. NO.: 2,41e-19 Length: 19723
Score: 457.50 Matches: 242
Percent Similarity: 33.1% Conservative: 46
Best Local Similarity: 27.8% Mismatches: 256
Query Match: 12.6% Indels: 325
DB: Gaps: 55

US-10-620-914-44 (1-1947) x US-10-084-846A-5 (1-19723)

QY 1945 AGTTGTCTCTTCGCGCCCTTCGCGGGCCATGTAGA----- 1907
DB 6222 SerThrSerGlyTTPArgPro-----GlyProCysArgGlyProSerValArgThrPro 6239
QY 1906 -----AGAGCTGTACATGTTGACGCGCTCAGTAC-----CCT 1871
DB 6240 AlaGlyArgLeuCys-----ProValSerAlaThrAlaAlaProThrPro 6254
QY 1870 GAGTGGCGCGCGATGACGCGACGTCGAGCGCCCTCTCGATCAGCTCGCGTAGG 1811
DB 6255 MetTrp-----SerSerAsnArgArgProProArgProProAlaProArgArg 6270
QY 1810 GCGGCTGAGGAGGCGGAGCGCCAGATGACGATGCCCGCGCGCAACCTGCTTGGCCA 1751
DB 6271 Ser-----ArgPro-----GluProCysSerArg 6278
QY 1750 GGCATCGGCCAGCTCGTTGGCCAGCGCATATCCAGCAGTCCACGTGGTCCATCAGAA 1691
DB 6279 CysProArgProProArg-----ArgArgSerValThrAlaArgProArgSerProSerGly 6297
QY 1690 TCACCTTGGTGTAGTGGCGCTTTGAGCTCTCCATGAAGAAGTTGGTGGAGACGGTCA 1631
DB 6298 Ser-----ArgArgAlaLeu-----ThrTrpProArgSer 6307
QY 1630 GGTGTGTCCA-----CCAGCCCACTCTTGGAGG-----TGGCGAAGG 1595
DB 6308 GlyThrProTrpProIleGlySerProArgTrpSerThrGlyArgProSerTrpArgPro 6327
QY 1594 -----CCGCTCGCGCAGGTAGTGGGC----- 1571
DB 6328 AlaGlyThrAsnTrpProProGlyCysAlaArgTrpProLeuArgAsnProArgProVal 6347
QY 1570 AGTTGTGCGCAGGAACCTTGGCGGTGAGGAGTGTAGTAGAAGTAGTGTCTGCTCGCA 1511
DB 6348 SerSerProGlyProArgCysArg-----MetProAlaValProSerGlyCysSer 6364
QY 1510 CGTGGAGTCTTCGCGCCACGCGCTCCATGGTGGCGCGGATGT----- 1469

6983 ThrTrpCysArgSerArgValCysArgArgSerCysValProGlyCysSerThrArg 7002
QY 1358 -----TCAAGTTCGTGAGCTGTCTTCA---ACAAGGCGTGTCTGT----- 1399
DB 7003 CysAlaSerArgSerArgArgCysAspSerProProArgProCysThrPheAlaAla 7022
QY 1400 -----GGTTGGCGGCGCGTGGCGGCAAGC 1426
DB 7023 ArgTrpTrpSerArgGlySerThrArgArgTrpTrpSerAlaAlaGlyTrpArgSerAla 7042
QY 1427 AGTACGCGCTGATCAAGCGGACGGATCC-----CCATTGAGAACTACATCG 1474
DB 7043 SerAlaGlyThrCysArgProTrpSerSerGlySerCysProProSerAlaThrSerSer 7062
QY 1475 CGCGCACCA-----TGGACGCGTGGCGGAGAACTGCGACGTGCGAAGCAGAACT 1525
DB 7063 IleThrProAsnAlaSerArgAsnThrArgArgArgThrAlaMetArgProGluArgThr 7082
QY 1526 ACTTCTACTACACTGCTCACCAGCAAGTTCTGCGCGCAACATGCGCCCACT----- 1579
DB 7083 AspSerValThrHisValSerArgGlyArgGlyGluArgAlaAlaProAlaGlyPro 7102
QY 1580 -----ACCTGCGGAGCGCGCTTCGCCA 1603
DB 7103 GluThrGlyLysProGluLysHisAlaSerThrIleValValGlyArgArgHisSerMet 7122
QY 1604 CCCTCA----- 1609
DB 7123 ProSerGlyValProLeuLysArgProAspLeuThrAlaLeuThrAlaArgCysGlyIle 7142
QY 1610 ---AGAGTGGGTGGTGACAACTGACCGCTCTCCACCACTTCTTCATGAGAGAGCTCA 1666
DB 7143 ArgArgSerTrpTrpTip-----AlaGlyLeuAlaArgGlyCys 7155
QY 1667 AAGCGCGCACCTACACCAAGGTGATCTGTATGACCACTG----- 1708
DB 7156 GlyArgLeuArgIleProArgProAsnAsnTrpSerArgTrpProIleGluLeuPheSer 7175
QY 1709 ---ACTGGCTGATA----- 1720
DB 7176 ProMetTrpTrpIleProSerArgArgProGluTrpArgArgSerGlyLeuSerSerGly 7195
QY 1721 -----TGCCCGTGGCCAAAGAGTGGCGCGCGCTGCGTGGCCAGC 1759
DB 7196 ArgProArgArgSerArgLeuProTrpGluThrAlaArgSerSerAlaTrpThrSer 7215
QY 1760 AGTTGGCGCGGCGCATCGTCATCTGGCGCTCGCCCTCCCTCAGCGCGCCCTACGCG 1819
DB 7216 ProThrSerSerArgThrProArgAlaGlyTrpProTrpProCysSerCysProValThr 7235
QY 1820 AGCTGATCCAGAAAGCGGCTTCGAGTGC----- 1849
DB 7236 Ser---SerAlaThrThrSerSerCysThrTrpAlaThrThrThrSerTrpThrAla 7254
QY 1850 -----GCTGCATCGCGCGCCCACTCAG----- 1873
DB 7255 SerSerThrSerTrpArgThrSerAlaAlaThrSerProProArgArgSerCysSerPro 7274
QY 1873 ----- 1873
DB 7275 GlySerProThrArgAlaGlySerAlaTrpProAsnLeuThrArgThrAlaGlyProArg 7294
QY 1874 -----GCTATGACGCGCGCTCAACATGTACA----- 1900
DB 7295 SerTrpArgSerProArgThrProGlyAlaThrTrpArgSerSerAlaCysThrProSer 7314
QY 1901 ---GCTCCTTCTACA-----TGGCGCGCGGAAAGCGCGCCAGAGGACAACT 1945
DB 7315 AlaArgProCysThrArgProTrpProAlaSerGlyArgProGlyGlyThrSer 7332
RESULT 6
US-10-084-846A-5

6365 ArgValThrGlyProSerGlyProGlyTrpAlaAlaAsnCysTrpLysProSerArgSer 6384
1468 AGTTCTCAATGG |||| |||| |||| |||| ||||
6385 SerProArgSerThrAsnTrpSerArgCysSerArgArgSerGlySerArgPro 6403
1441 TGATCAGCGCTACTGCTTGGCCCGCAGCCGCCCGCAACACACAGCA ---CGGCTTGT 1385
6404 ---GlyArgCysCysTrpArgAlaThrThrProArgSerThrAlaProArgProSer 6421
1384 TGAAGACACCA ---GCTCAGCAACTTGACGAACAGCCACACCA 1343
6422 SerArgCysSerSerGlySerProArgCysGlyGlyHisGlyAlaSerSerProPro 6441
1342 GCGGCTTGGCCCGTCTTCTCAAGT |||| |||| |||| |||| |||| 1316
6442 SerSerAlaThrArgSerValArgSerProProSerProProGlyArgProProThr 6461
1315 ---GGATGAGCA ---TGTGCTGTCCACAGAGCGGCTGCTCTCCCAATG 1271
6462 GlyProGlySerAlaValGlyArgCysCys ---AlaArgProArgGly 6477
1270 TGGCGCTTGGCGGCTTGACGCTTGGCCAGTCCACAGCA --- 1223
6478 ValAlaArgTrpArgTrpSer ---ValCysProSerThrArgProProSerGlySer 6495
1222 ---CGGCCAGGCACTGCAGCACCCAGCACAGCT 1193
6496 ArgAlaMetThrArgTrpSerArgProSerArgProArgArgProProSerProVal 6515
1192 TGCCCATGCCCGCTGTAGTACAGCGCTGTGAAGTACAGAGCGCTTGACACAGA 1133
6516 ThrProAlaArgSerArgSerAlaAlaGlyProThrArgGly ---TrpSerTyr 6534
1132 AGTTGTGC ---TGCTTGGCAGAGG |||| |||| |||| |||| 1106
6535 GlyGlySerProProThrTrpProSerThrAlaAlaArgTrpTrpProCysSerThrGly 6554
1105 ---CGGCCAGCT ---TCTTCTCGTACAGCT 1082
6555 CysAlaArgProProThrSerValProAlaHisArgThrArgSerThrArgArgArg 6574
1081 CCTCAATGCCCGGTGCAGCCCT ---CGCCGAACAGCTGCCACA 1040
6575 SerArgThrArgGlyProArgProProThrAlaAlaGlyProThrArgGlyCysAlaThr 6594
1039 CGT ---CCTCAACTCCA ---GCTGTGAATGCCACCTTCT 1004
6595 ArgCysGlySerProProArgSerProArgArgProArgThrAlaThrGlyArgSerSer 6614
1003 TCAGCTCCAGAACGCCGACTGCG --- 980
6615 SerCysProArgThrProTrpSerArgThrArgSerThrArgTrpProAsnAlaAla 6634
979 ---CGGGTTGCAGT ---CCACCG 962
6635 TrpArgThrCysSerSerArgArgTyGlyAlaThrSerProArgProGlyProSer 6654
961 ACACACTGCCCGGCCCT ---GCACCAGGTTTCAGGCAATGACGCCCGC 908
6655 AlaProProTrpAlaProProThrAlaThrAlaAlaTrpThrGlyArgSerArg 6674
907 TAGTCAGGTCAGCACCGTGTCTTGGGTTGATCTCCATCACT --- 863
6675 ThrGlyThrSerLysSerCys ---ArgProThrProGlySerThrAlaArg 6690
862 ---CATATCCGCTCGGGTCTTCCACAGCTCGTGTACAGGAAGTGGCG 812
6691 CysGlyValProSerProGlyArgGluArgProSerAlaAlaThrThrSerThrProThr 6710
811 GGAACATGGCG ---GCGCTCCAGCGCTCTCTGTCAGGCGGTGGC ---CAACGC 761

6711 ArgCysTrpAlaArgProAlaAlaSerArgAlaAlaThrCysGlyCysGlyThrAlaArg 6730
760 TGGCGAGCGCGCAATCCACACAGCTAGTAGGGCGCGCAGCC ---ACGCACGT 710
6731 TrpThr ---ThrThrAlaAlaArgThrArgAlaAlaThrProSerThrAlaTrp 6747
709 AGGGATCGAACCTGGGTGT ---TCTGCTCCACACAGCGCT 671
6748 ArgSerSerArgProProCysTrpArgSerArgSerTrpArgProAlaPro ---ArgAla 6766
670 CCAGCTTCTGCTCCAGGTAGCGCGCTCGGGCCGATGTCATGTTGTCGATCGA 611
6767 LysSerAlaAlaPro ---SerArgThrArgProThrArgCysArg 6780
610 AGATCGATCGCCAGAGAAACCGCGCACAGGCGCATCTGGCGCAGG ---GCAGGT 557
6781 ArgValSerGlyArgSerArgSerSerAlaArgAlaArgTrpCysGlyTrpProProGly 6800
556 CGTACTTCCGCTCACGTAGAAAGTGGCAACCGCCACCGCCGCTTGGGACAGTACG 497
6801 ArgSer ---ArgThrProProThrArgThrProProGlySer 6813
496 AGC ---AAGCTGTGCTGATGACGTGTGCGAACGGTGAATCATCTGAGCAGTAGG 443
6814 SerMetProArgProGlyArgProArg ---ArgThrSerProValTrpArg 6829
442 AGAAGGTGATGAGCGTCCGCTGCGCTCAGGGGCGCAAAATGGCAAGCTCGGCTCCA 383
6830 Arg ---GlyArgCysTrpThrArgAlaSerThrGlySerAsnArgPro --- 6844
382 CGACTGGACATCTTCCAGCCCTTGGCTTCTTCTTGGCCACCTCGCACAGCG 323
6845 ---ThrProAlaTrpSerProGlyGlyTrpPro --- 6854
322 AGTGGCAGCTGCACACAGTGTGACTTGAACCTCGCCAGT --- 278
6855 ---ArgTrpAlaTrpProArgProValSerIleGlyAlaSer 6867
277 ---CGATGTAATCAGCCATCATATCGA 254
6868 SerAspCysProProValSerValTyCysThrLeuArgCysSerArgProThrProArg 6887
253 CATCTCCAGTGCACACCGCTCAACCCAGATGAGGTTCGAGCGCTCGGCCA --- 197
6888 ProGlyProArgCysTrpThrProCysArgSerArg ---ArgProPro 6902
196 ---GGC 194
6903 SerTrpAlaSerArgSerSerAlaTrpSerCysThrSerThrArgSerProSerThrAla 6922
193 GGGCAGCAAGGAGCGCGCTGGGGCCGTAGAGCTCTCCAGCGCAGCAG --- 143
6923 ArgHisArgArgArgArgSerArgSerIleProAlaSerProThrProCys 6942
142 ---CGTGATCATCGCCCTTCTTGTCTCCGAACACCATATGCG 104
6943 ThrProTrpSerArgThrGlyArgAspAlaArgTrpArgAlaCysAlaAlaCysAlaThr 6962
103 GCAGAACGTCAGGTATCTTCTTCTGCTGAGCTTGAGCTTCTCCAGGAGAGTTCT 44
6963 ArgTrpSerSerArgProArgTrpProThrSerArgAlaAlaThrArgThrArg 6982
43 TCTTGTGTAGTCGACGCGCGCT 17
6983 ThrTrpCysArgSerArgValCysArg 6991

RESULT 7

US-10-084-846A-7
; Sequence 7, Application US/10084846A
; Publication No. US2004006026A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUHLENWEG, AGNES

; APPLICANT: TREFZER, AXEL		; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES	
; FILE REFERENCE: 1974-005		; CURRENT APPLICATION NUMBER: US/10/084, 846A	
; CURRENT FILING DATE: 2003-02-25		; PRIOR FILING DATE: 2001-08-24	
; PRIOR FILING DATE: 2001-08-24		; PRIOR FILING DATE: 2001-02-25	
; NUMBER OF SEQ ID NOS: 120		; SOFTWARE: PatentIn Ver. 3.2	
; SEQ ID NO 7		; LENGTH: 19652	
; TYPE: PRT		; ORGANISM: Streptomyces viridochromogenes	
; FEATURE:		; OTHER INFORMATION: Protein 2: amino acid sequence encoded by coding strand 2.	
; OTHER INFORMATION: Start codon: gat, Start position: nucleotide 2.		US-10-084-846A-7	
Alignment Scores:		Pred. No.: 2, 77e-19 Length: 19652	
Score: 456.50 Matches: 232		Percent Similarity: 36.3% Conservative: 38	
Best Local Similarity: 31.2% Mismatches: 247		Query Match: 12.6% Indels: 227	
DB:		Gaps: 50	
US-10-620-914-44 (1-1947) x US-10-084-846A-7 (1-19652)			
QY	1945	AGTTGCTCTCTTGGCGCCCTTCGCGCGGCATGTAGAGAGCTGTACATGTTGACGC	1886
Db	1918	AlaCysProAlaThrLeuProPheProPro	1928
QY	1885	GGTCCATGTAGCCCTGAGTGG	1865
Db	1929	GlyProAlaHisProLeuTriLeuAlaProLeuArgThrProAlaArgProAlaGlyLeu	1948
QY	1864	-----CGCGCGCGA	1856
Db	1949	LeuProProLeuAlaArgProArgProAlaProArgValProLeuProValArgProGly	1968
QY	1855	TGCAGCGCAGTCGAGCCCGC---CCTTCTGATCAGCTCGG	1817
Db	1969	-----ThrArgArgThrProArgProSerAlaArgSerArgArgProGlnArgSer	1986
QY	1816	CCTAGGGCGGTGAGGGAGCGGAGCGGCAGATCAGTCGCGCCCGCGCAACTGCT	1757
Db	1987	ArgAlaAlaGlyAlaProSerArgThrAlaPro---AspAlaArgThrProProCysAla	2005
QY	1756	TGCCAGGACCTCGGCCAGCTGTTGGCCAGCGGCATATCCAGCCAGTCACGTGTCCA	1697
Db	2006	SerProSerCysArgArgAlaArg---ProArgGlyArgPro---ProArgProPro	2022
QY	1696	TCAGATACCTTGGTGTAGGTGCGCGCTTTCAGCTCTCCATGAGAAAGTTGTTGAGA	1637
Db	2023	CysTrpAlaProProThrArgProCysAlaGlyAlaProGlyValArgGly---Arg	2040
QY	1636	CGGTACAGTTGT	1625
Db	2041	ProAlaGlyCysGlyAlaAspArgProValArgProProCysGlyThrProTrpSerPro	2060
QY	1624	CCACACCGCCTCTTGGAGTGGCGAGCGCCCTCGCGCAGTAGTGGGCGACTGT	1565
Db	2061	ProArgArgArgAlaArgThrProProProAlaArgAlaGly---GlySerAla	2078
QY	1564	---CGCGCAGGAACCTTGGCGGTGAGGAGT---TGAGTAGAAGTAGTCTGCTTGC	1514
Db	2079	ProGlyCysGlyThrProArgArgGlySerProTipAlaThrArgSerGlyThrAlaArg	2098
QY	1513	GCAGTGCAGGATGCTCCGCCA-----CGCGCTCCATGTTGCGCGCATGT-----AGT	1466

Db	2099	GlyArg---SerAlaProProThrArgArgProTrpSerGlyGlySerGlyAspSer	2117
QY	1465	TCTCAATGGGATGCGCTCGCTTATCAGCGCGTACTGCTTGGCCGCGCACCGCCGC	1406
Db	2118	SerArgSerGlyProArgPro-----ArgArgAlaCysProProArgArgPro	2133
QY	1405	CGAACACACACGCGCTTGTGAGA-----GCACACAGGC	1370
Db	2134	ArgAlaProGlySerProAlaAlaArgArgArgTrpProArgArgArgProAlaProSer	2153
QY	1369	TCACGAACCTTGCAGAACACACACAGCGCTGGGCCCTTCTTCCAGAGTGGATGA	1310
Db	2154	SerArgArgThrArg-----ProAlaGlyArgGlyArgAlaAlaValSerArgAla	2170
QY	1309	GCATGTTGCTGT-----CCACA	1292
Db	2171	AlaHisCysAlaSerArgSerProGlySerSerAlaArgTrpProAspArgProPro	2190
QY	1291	GACGCGCTGCTCT---CCATTGTGGCG-----CCTTGGCGAGCGCTTGACGG	1244
Db	2191	SerGlySerAlaProArgArgLeuGlyAlaArgValArgArgTrpProGlySer-----	2208
QY	1243	TCTTCCCGCTCCAGCACACCGCCAGGCACTGCAGACCCAGC-----ACAGCT	1193
Db	2209	SerCysArgAlaProSerArgArgArgGlySerSerGlyAlaGlyArgGlyArgSerGly	2228
QY	1192	TGCCATGCGCCCTGTGTAGTACAGGCGCTGCTGGAAGTACAGAGGCGCTTGACACGA	1133
Db	2229	CysArg---ArgProGlySer-----ArgAlaValArgProGlyGlyAlaCysArgArg	2245
QY	1132	AGTTGCTGCTGTTGGCACAGGAAGCGCGCAGCTTCTCTCGTACAGCTCCTCAATGC	1073
Db	2246	AlaProGlySerProAlaArgProProArgProAlaGlyThrProArgProProArgGly	2265
QY	1072	GCGGGTGACACCCCTCGC---CGAACAGCTGCCACAGCTCTCAAATCCAGCTGCTGAA	1016
Db	2266	ArgGlyArgGlyProArgAlaArgAlaGlyArgSerThrArgSerAlaAlaArg	2285
QY	1015	TGCCACCTTCTTACGCTCCAGAGCGCGACTGCGCGGGTTCAGTCCACGACCA	956
Db	2286	SerArgProGlyArgProProArgSerPro-----GlyGlyAlaAlaGlyArgProArg	2303
QY	955	CCTGCGCGCCCTTCGACACAGCAGTTTCAGGCGCATTCAGCGCGCTAGTCAGGGTCA	896
Db	2304	ProGlyArgAlaAlaValAlaHisArgArgGlyArgProAlaArgSerValArgGlySer	2323
QY	895	GCACCGTCTCTTGGGGTGTATCTCCATCATCTCCATATCCAGCTAGTAGGGGCGCGCAG	836
Db	2324	ProPro-----ArgProProAlaProValProValArgProAlaProArg	2338
QY	835	ACTGCTGTACAGGAAGTGGCGGGAACATGGGCGCGCTCCACGCGCTCTCTGTGCA	776
Db	2339	ThrAlaCysAla-----AlaGlyLeuProProAlaProProArg---ProAlaAla	2354
QY	775	GGCGCTGGCCAAACGCTGGCGCGCCAAATCCACAGCTAGTAGGGGCGCGCAGCCAG	716
Db	2355	GlyArg-----AlaSerSerAlaProArgTrpProArgArgProGlyVal	2370
QY	715	GCACGTAGGGGATCGAACCTCGGTGTTCTGTCTCCACACGCGCTCCAGCTTCTGTCCA	656
Db	2371	ProArgGlyProProProProAlaArgAlaAla-----ArgArgProValArgTrpSer	2388
QY	655	GGTAGCGCGCGCTCGGGCGCATGT-----CAATGTTGT-----CGATGCCAAGA	608
Db	2389	GlyAlaArgAlaCysProArgCysValaArgCysCysProValGlnProArgArg	2408
QY	607	TCGATCCCGA-----AGAAACGCGCGCACCGCATCTGCGCGAGGG	563
Db	2409	SerAlaGlyArgArgProProSerGlyArgArgSerAlaAlaArgProArgGlyAlaGly	2428
QY	562	GCAGGTGCTACTTGGCGCTCACTAGAGTGGCAACGCCACCGAGCGCT---CTTGGG	506
Db	2429	AlaAlaGlyThr-----SerArgArgAlaProGlyArgProSerGly	2443

505	QY	ACAGGTACGAGCAAGCCTGGTCGATGACGCTTGTTGGAACGCTGGAATCATCGTGACGAGT	446
2444	Db	 ThArgProSerProProProPro	2451
445	QY	AGGAGAAGGTGATGACGCGTCGCGGTGGCCCTCAGGGGGCGCAAAATTGGCAAGCGTCGCGCT	386
2452	Db	 ::: -----GlyAlaAlaCysProArgGlyProThrAlaGlyProProGlyPro	2466
385	QY	CCA---CGACTGGA-----CATCTTCACGCCCTTGGCTTCGCGCTTCTCTCTGG	338
2467	Db	 ProAlaArgAlaGlyProValAlaHisGlySerValProGlyProVal-----	2483
337	QY	CCACTCGCACAGCGATGGGCACAGGTCGACCAAGTAGATGGACTTGAACTTCGCCAGGT	278
2484	Db	 -----ArgProArg	2487
277	QY	CGATGTAATCAGGCATCATATCGACATTTCTCCCGCAGTGCACCAAGCCAGGTCACCCAGA	218
2488	Db	 ::: -----ArgValArgAspArgProPro-----ProAlaProProArg	2500
217	QY	TGAGGTTTCAGCCCTCGGCCAGCGGGCGGCACAAAGCAGCGGCTGGGGCCCGTAGAAGC	158
2501	Db	 ::: -----GluAlaAlaArgProGly-----ProProProAlaHisSer	2513
157	QY	TCCTCCAGGGCAGCAGGT---GATCATCG---CCCTCTGTGCTGCCGAACCATATGGC	104
2514	Db	 ::: AlaProSerArgProArgProSerHisArgValProAlaSerCys-----	2528
103	QY	GCAGAACGGTCAGGTCATCTCTCATGCTGCTGAGCTTGAGCTTCTCCAGGGAGGAAGTTCT	44
2529	Db	 -----HisPro-----GlyGlyArgGlyLeAsp	2536
43	QY	TCCTGGTGATGC	32
2537	Db	 ::: GlyTrpCysAla	2540

RESULT 8

```

US-10-084-846A-8
; Sequence 8, Application US/10084846A
; Publication NO. US20040006026A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUHLENWEG, AGNES
; APPLICANT: TREPFER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084, 846A
; CURRENT FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 8
; LENGTH: 19608
; TYPE: PRT
; ORGANISM: Streptomyces viridochromogenes
; FEATURE:
; OTHER INFORMATION: Protein 3: amino acid sequence encoded by coding strand 1.
; OTHER INFORMATION: Start codon: atc, Start position: nucleotide 3.
US-10-084-846A-8

```

Alignment Scores:	1.23e-18	Length:	13608
Pred. No.:		Matches:	205
Score:	446.00	Conservative:	57
Percent Similarity:	37.5%	Mismatches:	248
Best Local Similarity:	29.4%	Indels:	188
Query Match:	12.3%	Gaps:	34
DB:	4		

US-10-620-914-44 (1-1947) x US-10-084-846A-8 (1-19608)

Qy	17	ACGCCCGCCG-----CTGCGAGCTACACCAAGAAGAACTTCTCCCTGGAGAACTCA	67
Db	5039	ThrAlaGlyArgThrProLeuArgArgAlaProArgGlyThrHisPro-----	5054
Qy	68	AGCTCAGCAGCATGAAGGATGACCTGACCGTTCTGGCCCATATGTGTTCCGCGAGCAAGA	127
Db	5055	-----AlaSerGlySerArgArgArg	5061
Qy	128	AGGCGCATGATCACGCTGCTCGCTGGAGAGCTTCTACGGGCCCCACGGCCGCTGCCTTTG	187
Db	5062	ArgSerSerGlnValArgLeuGlyGlnTyrAlaGlySerHisAlaPro-----	5078
Qy	188	CTGCCCGCCCTGGCCGAGCGCTCGAACTCATCTGGGTTGCACCTGGGTGTGGCA-----	241
Db	5079	LeuGlnAspGlyArgGlyAlaArgHisSerArgSerGlyCysTyrIleValAlaHisGlu	5098
Qy	242	-----CTGGGGAGAATGTCTGATATGATGGCTG	268
Db	5099	GluGlyLeuGluIleProLeuGluArgArgAspTyrThrArgArgSerIleSerGlnLeu	5118
Qy	269	-----ATTACATCGACCTGG	283
Db	5119	ProValProAspValHisPheTyrProGlnCysAlaAlaThrAsnLeuAsnTyrGlnTyr	5138
Qy	284	CGAAGTTCAAAGTCCATCTACGTGGTGCAGCTTGCACCTGCGCTGTCGAGGTGGCCAAGA	343
Db	5139	GlyAlaProCysProIlyslsCysSerSerAlaValGlyArg-----ArgAspProArg	5156
Qy	344	AGAAGCGAAGCCAAGGCGCTGGAAGAATGTCCAGGTGCTGG-----AGGCCGAGCTT	397
Db	5157	ArgProProCysSerAlaArgProAlaCysProSerTyrTyrAsnGlyArgProSerArg	5176
Qy	398	GCCAAATTTGGCCCCCTGAGGCGACCGCGACGCTCATCCTTCTCTCTACTCGCTCAGA	457
Db	5177	AlaThrThrSerAlaSerArgSerProProArgAlaAlaProSerSerThrSerSerAla	5196
Qy	458	TGATTCCACCGTTCCACAAAGTCAACGAGTTCGCTGCTGCTGCTGCCAAGACGGCC	517
Db	5197	ArgTyrArgArgSerThrProAlaValThrAlaArgSerArgThrAlaCys-----	5213
Qy	518	TGTTGGCGCTTCCGACTTCTACGTGAGCGGCAAGTACGACCTGCCCTTGGCCAGATGC	577
Db	5213	-----CysCysAlaGlyAlaArgArg	5219
Qy	578	CCTGTTGCGCGCGTTCTTCTGCGCATCGATCTTCGACATCGACAACTTGACATCGGCC	637
Db	5220	ThrGlyArg-----SerThrGlyArgArgSerSerValPro-----ThrSerAla	5234
Qy	638	CGG-----AGCGCGCGCGCTACCTGGAGCAGAAGCTGGAGCGCGTGGGAGCAGAGA	691
Db	5235	ProGlyArgSerThrAlaThrThr-----SerThrTyrCysSerArgThr	5249
Qy	692	CCCAGGGTTTCGATCCCTACGTGCGCTGCTGCGCGCCCTTACTACGTGTGGATTTGCC	751
Db	5250	ArgSerSerArgAlaProThr-----SerProArgAlaSerArgSerSerAla	5265
Qy	752	GCGTGCCAGCGTTGGCCACGCCCTGACGAGGACGGGTGGAGCGCGGCCCATGTTC	811
Db	5266	CysCysSerThrAlaThrAlaArgLeuGlyArgSerGlyTyrProThrProValSer	5285
Qy	812	CGCCCACTTCTGTACAGCAGTGTGGGAGGACCCCGACCGGATATGGAGGTGATGG	871
Db	5286	-----TrpSerAlaArgSerSerThrThrSer-----	5294
Qy	872	AGATCAACCCCAAGCACACGGTGTGCTGCTGACTAGCGGGCGGTGCAATGCCCTGAACC	931
Db	5295	---SerThrProArgAla-----	5299
Qy	932	TGCTGTGTCAGGGGCGCGCCACAGGTGTGTGTGGTGCATGTCAACCCCGCAGTGTGGCGC	991
Db	5300	-----GlyProAla-----SerSerProSerThrArgThr	5309

QY	539	ACGTGAGCGGCAAGTACGACCTGCGCCCTTGC	CGCCAGATGCGCCCTGTGCGCGCGTTCTTCT	598
DB	8603			:::
		ThrThr-----	ProAspArgThrThrAspAla	8611
QY	599	-----GGCGATCGATCTTCGACATCGACA	CATTGCATCGCGCCCGGAGCGCC	646
DB	8612			
		TrpArgArgPheArgAspArggthrCysSerPro	ThrThrProArgTrpArgArgCysAla	8631
QY	647	CGCGCT---ACCTGGAGCAGAAGCTGGAGCGC		679
DB	8632			
		ArgProAspProTtpPheArgSerArgSerAla	AlaThrTrpTtpCysThrThrSer	8651
QY	680	GGGAGCAGAACACCCAGGTTTCGATCCCTAC	GTGCGCGTGC	721
DB	8652			
		GlySerThrValArgAlaArgThrProThrAla	ValGlyGlyProSerSerSer	8671
QY	721	-----	-----	721
DB	8672	SerThrProThrArgProHisArgArgProGlu	ArgCysAlaAlaGluArgGlnAla	8691
QY	722	-----	-----	733
DB	8692	AspIleGluArgGlnValArgGluSerPheGly	LeuCysSerProProArgThrGly	8711
QY	734	-----ACTAGTGTGATTTGGCGCTGCCAGC	GTGTGGCCACGCCC---TGCACGAGG	784
DB	8712	ArgAspIleThrCysAlaTtpCysProTtpPro	GlyArgCysAlaProArgAspThrArg	8731
QY	785	AGCGCTGG-----	-----AGCGCGCCCATGTTCCCGCCCA	817
DB	8732	SerGlyTrpArgAlaArgArgArgAsnAlaGly	SerProArgProAlaCysArgSer	8751
QY	818	CTTCTCTGTACACGCAGTCTGTGGGAGACCC	CGGATATGGAGGTGATGGAGATCA	877
DB	8752	ArgCysSerThrArgArggThrTtpLysAlaPro	Gly-----TtpArgSer	8766
QY	878	ACCCAAAGCACCGTGTCTGACCTGACTAGC	GGCGGCTGCANATGCCCTGCAACTCTCTGG	937
DB	8767	-----MetCysArgPro-----	-----	8770
QY	938	TGC-----AGGGGCGGCCACAGTGG---	TGTCGTGGACTGCACAGAGAACTGGCGC	1105
DB	8771	CysThrHisArgArgSerProValArgGlyArg	CysArgCysThrArgSerProGlySer	8790
QY	986	CGGCGCTTCGAGCTCAAGAAGTGGCCATT	CAGCAGCTGGAGTTTGAGGAGCTGTGC	1045
DB	8791	Arg-----	-----CysSerArgSerThrThr	8797
QY	1046	AGCTGTTTCGGCAGGGCGTGCACCCCGGCAT	TGATGAGGAGCTGTACGAGAGAACTGGCGC	1105
DB	8798	SerThrProAlaThrCysAlaThrSerGlyArg	SerArgSerThrArgCysSerAlaAla	8817
QY	1106	CTTCTCTGTGCGAACCAGCCACAACCTTCGT	GTCGCAAGCGCCTCTGGTATCTCCAGCAGC	1165
DB	8818	MetThrThrArgAlaSerAlaThrThrGlyPro	-----ThrTtpTtpCysThrThr	8835
QY	1166	GCCTGTACTACGAGGGCGCATGGCAAGCTG	TCTGGGTCGTCAGTGCCTGGCGCTGG	1225
DB	8836	SerTtpProSerArgAlaHisTtpSerProArg	SerGlyAlaCysArgAla-----	8852
QY	1226	TGCTGGACTTGGGCAGACAGCTCAAGCGCT	CGCCCAACGCGCCCAAA-----	1273
DB	8853	-----SerMetCysArgProGlySerSerAla	ProTtpArgProSerArgGlySerThr	8870
QY	1274	TGG-----	-----AGGAGCAGCGCCTCTGTGGGACAGCA	1306
DB	8871	TrpSerArgThrArgCysArgAlaSerArgSer	ThrAlaSerSerGlyAlaAlaThr	8890
QY	1307	TGCTCATCCACTTCGTGAAGAACGGGCCCA	-----AGCGCTGG---	1345
DB	8891	GlySerSerThr-----ArgTtpThrProHis	ArgThrTrpArgCysArgArgTrpAla	8908

1346	Qy	-----TGTCGCTGTTGCTCAAGTTCTGTGAGCCTCGTCTCTTCTCAACAAGCCGG-----	1399
8909	Db	ThrArgCysGlyCysProCys-----AlaThrCysProThrThrAlaProArgAla8925	
1394	Qy	-----TGCTGTGTTGTCGGCGCGCGGTGC1417	
8926	Db	ArgThrArgGlySerTrpGlyArgSerGlyAlaAspAlaCysAlaSerSerGlyAlaThr8945	
1418	Qy	-----CGGCAAGCAGTAGTCGCCTGATCAAGCGGACGGCATCCCA---TTGAGAACT1468	
8946	Db	ArgProArgAlaSerSerAlaProThrTyArgArgCysGlyThrProSerThrArgArg8965	
1469	Qy	ACATCGGCGCA-----CCATGGACGGCGTGGCGG1498	
8966	Db	CysGlyArgAlaArgArgTrpCysSerArgArgArgAsnArgTrpArgSerTrpGlu8985	
1499	Qy	AGAACTCGCACGTGCGCAAGCAGAACTACTTCTACTACAACATGCTCTCACGGCAAGTTC1558	
8986	Db	ArgCysArg-----ProAlaSerGly8992	
1559	Qy	TGGCGGACAACTGCCCCACCTACTCTGCGAGGCGGCTTTCGCCACCTCAAGAGTGGCG1618	
8993	Db	CysCysAlaThrAla-----ArgTrpSerSerSerArgThrAla9006	
1619	Qy	-----TGTGGACAAC---TGACCGTCTCCACCAACTTCTTCATGGAGGAGCTCAAAG1669	
9007	Db	ThrCysTrpCysThrThrAlaAlaProThrAlaThrAlaSerTrpArgGlyPheArg9026	
1670	Qy	CGCGCACTTACACCAAGGTGATTCTGATGACCACTGGACTGGCTGGATATGCCCGTGG1729	
9027	Db	SerCysProSerProSer-----ThrThr-----9034	
1730	Qy	CCAACGAGCTGGCGGAGTGCTTGGCCAACAGAGTTGCGCGGCGGCATCGTCATCTGGC1789	
9035	Db	-----ThrArgSerAlaAlaValSerIleProPro9045	
1790	Qy	GCTCCGCT-----CCCTCAGCCGCCCTACGCCGAGCTGATCCAGAGGCGGGCTTCG1843	
9046	Db	AlaProProCysProAlaLeuArgProArgProArgSerArgArgCysAla9065	
1844	Qy	ACGTGC-----GCTGCATCCGCGCGCCACTCAGGGCTACATGGACCGCGTCAACA1894	
9066	Db	GlyCysCysSerThrThrAlaThrAlaGlyArgArgArgAsnCysGlyThrAlaSerSer9085	
1895	Qy	-----TGTCACAGCTCCTTCTACATGGCGCGCGCGGA1924	
9086	Db	GlyHisLeuArgProProArgSerProGlyCys-----ProAlaSerTrpArgThrAla9103	
1925	Qy	AGGGCCCAAGAGGACAAC1945	
9104	Db	GlyCysArgArgArgThrSer9110	

; TYPE: PRT
 ; ORGANISM: Streptomyces viridochromogenes
 ; FEATURE:
 ; OTHER INFORMATION: Protein 3: amino acid sequence encoded by coding strand 1.
 ; OTHER INFORMATION: Start codon: atc, Start position: nucleotide 3.
 US-10-084-846A-8

Alignment Scores:
 Pred. No.: 5,05e-18 Length: 19608
 Score: 436.00 Matches: 217
 Percent Similarity: 35.0% Conservative: 40
 Best Local Similarity: 29.6% Mismatches: 237
 Query Match: 12.0% Indels: 240
 DB: 4 Gaps: 44

US-10-620-914-44 (1-1947) x US-10-084-846A-8 (1-19608)

```

QY 1939 CTTCTTGGCGCCCTTCGGCGGCCATGTAGAGGAGCTGTACATGTTGACGGGTCCA 1880
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1879 TGTAGCCCTGAGTGGCGGGGGATGACAGCCAGCTCGAAGCCCGCTTCTGATCAGCT 1820
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 10550 -----ProArgArgProSerProThrAlaArgProProArgGlyGlySerAla 10567
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1819 CGGCGTAGGGCGGCTGAGGAGGGGAGCCAGATGATCGCCGCCCGCCGCAACCT 1760
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 10568 AlaArgPro-----ArgArgAlaValProAlaArgArgProAlaValSer 10585
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1759 GCTTGGCCAGGCACTCGGCCAGCTGTTGGCCAGCGGATATCCAGCCAGTCCACGTGGT 1700
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 10586 ProProProGlyGlySerGlyGlyArg-----ThrAlaProGlyAlaAlaProProArg 10603
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1699 CCATCAGAATCACCTTGTGTAGTGTGCGCTTTGAGCTCTCCATGATGAAGAAGTTGGTG 1640
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 10604 ProSerArg---ProTriProGln----- 10610
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1639 AGACGCTCAGGTTGTCCACGAGCCACTCTTGAGGTGGGAGGCCCGCTCGC---GCA 1583
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 10611 GlyArgThrGlySer-----GlyTriArgArgArgAspArgTyGly 10624
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1582 GGTAGTGGGCGAGTGTGCGCAGGAACCTTCCGCTGAGGCGAGTGTGTAGTAGAGTAGT 1523
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 10625 GlyArgArgSerAlaCysArgSerGly-----ArgGlyThr 10636
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1522 TCTGCTTGGCAGCGTGGAGTTCTCCGCCAGCCGCTCCATGCTGCGCGGATGTAGTTCT 1463
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 10637 Ala-----AlaGluSerProProArg-----CysArgSerCysSerSer 10649
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1462 CAATGGGATGCGCTCGCTTGATCAGCGGTACT-----GCT 1424
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 10650 ---TriArgCysArgGlyAlaProAlaGlyArgSerThrGlyGlyAlaProAlaGlySer 10668
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1423 TSCCGGCGAGCGCGCGCGGACACAGCAGCGCTTGTGAGAGCAGCCAGGCTCACGA 1364
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 10669 CysAlaAlaProGlyArgArgProAlaAlaLeuPro-----AlaPro----- 10682
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1363 ACTTGACGAACACCCAGCAGCGCTTGGGCCGCTTCTTCCAGAGTGGATGAGCATGT 1304
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 10683 -----ProProValProAlaAlaProAlaPro----- 10691
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1303 TGCTGTCCACAGAGCGGCTGCTCTCTCCATGTGTGGCGGCTTGGGAGCGCT----- 1250
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 10692 -----ProProArgProAlaAlaProPro-----ArgArgTriArgAlaAlaAlaPro 10707
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1249 -----TGACGGTCTTGGCCAGTCCAGCA---CCACGG----- 1220
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 10708 ValArgArgArgArgCysProArgProSerCysProArgArgCysThrArgValCys 10727
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1219 CCAGGCACTGCACACCCAGCAGCTTGCCTCCATGCGCCCTGGTAGTACAGCCGCTGT 1160
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 10728 ProGly---SerAlaProAlaAlaThrGlyPro---ArgProGlyArgArgCysArgAla 10745
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

QY 1159 GGAAGTACCAGAGCGCTTGGACCAGAGTGTGGCTGTTCGACAGAAAGGCGCCCA 1100
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 10746 ArgSer-----AlaTriAlaAlaAlaAlaArgArgPro 10756
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1099 GCTTCTTCTCGTACAGCTCTCAATGCGCGGTGCACGCCCTCGCGAAACAGTGCACA 1040
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 10757 Ala-----AlaAlaProProArgProValArgGly 10766
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1039 CGTCTCAAACTCCAGCTGCTGAATGCCACCTTCTTCAGCTCCAGAAAGCGCCAGTCG 980
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 10767 ProProArgThrGlySerGlyArgGlyProThrArgArgArgThrSerArgSerThr 10786
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 979 CGGGTTGCACTCCACCGACACACCTGCGCGGCCCTCCACACAGAGTTACGGCAT 920
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 10787 AlaGly-----ProProArgProGlyArg-----AlaProAlaArgSerProArg 10801
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 919 TGCAGCCGCGCTAGTACAGGTGACACCGTGT---CCTTGGGGTTGATCTCCA---TCA 866
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 10802 CysArgThrArgAlaThrHisArgArgSerAlaArgProGlyAlaArgProValArg 10821
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 865 CTTCCATATCCGCTCGGGTCTCCACGACTCGGTGTACAGGAAGTGGCGCGGAACA 806
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 10822 ProArgArgProCysArgAlaProProGlyArgAlaAlaSerGlyArgArgSerPro 10841
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 805 TGGCGCGCGCTCCACGCGCTCTCTGTCAGGGCGTGGCCAAACCTGGGAGCGGCCAA 746
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 10842 ArgCysHisSerCysAlaAlaProArgSerThrSerSerArgArgAlaGlyArg 10861
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 745 TCCACAGTGTAGGGCGCGCGAGCCAGCCAGTACGAGTGGGATGAACCTGGGTGCT 686
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 10862 ArgArgArgSerArgAlaThrAlaSerAlaAlaArgThrGlyAlaAlaAlaAlaAla 10881
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 685 GCTCCACACGCGCTCCAGCTTCTGCTCAGGTAGGCGCGCGCTCGGGCGCGATGTCAA 626
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 10882 AspArgThrAlaAlaValCysThrTriProAlaHisArgThrProAlaGlyArgPhePro 10901
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 625 TGTGTGATGTCGAAGATCGATCGCCAGAAACCGCGCAGCAGG----- 578
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 10902 SerCysProAlaAlaArgAlaProGlyProAlaSerGlyArgSerArgArgProArgVal 10921
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 577 -----GCATCTGGCGCA 566
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 10922 ProAlaThrAlaGlyThrProProGlySerArgValProArgArgAlaAlaProArgPro 10941
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 565 GGGCAGGTCTGCTTCTGCGC-----TCACGTAGA-----AGTCGCAAGC 524
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 10942 GlyAlaProCysThrCysAsnValProProSerSerArgArgThrSerSerSerGluArg 10961
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 523 CCACCA-----GGCGCTCTTGGGACAGGT 500
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 10962 LeuProAspLeuProPheAspValCysLeuAlaLeuGlyGlyAlaAlaAlaAlaPro 10981
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 499 ACGAGCAAGCTGTGTCGATGAGTGTGGAACGGTGGATCATCTGAGCGAGTAGGAGA 440
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 10982 AlaValArgProGlyArg-----ArgThrArg 10990
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 439 AGGTGATCAGGTCTCGGTGC-----CCTCAGGGGCGCAAAATTCGCAAG 395
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 10991 ArgTriSerAlaAlaAsnCysArgTriArgSerGlyProAsnArgGlyArgAlaGlyArg 11010
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 394 CGTGGCGCTCCA---CGACCTGGACATTTCTTCCAGCCCTTGGCTTCGCT----- 347
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 11011 ArgAlaProProArgArgProGly-----ProGlyProGluProArgValArg 11026
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 346 -----TCTTCTGGCCACCTCGCAGCGAGTGCACAGGT----- 311
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 11027 SerArgAlaProSerProProArgArgArgArgGlyAlaGlyAlaAlaProGluProPro 11046
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 310 -----CGACCACCTAGATGACT----- 293
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 11047 ProGlyIleArgCysAlaValArgArgArgProGlyArgCysSerAspCysArgProCys 11066
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 292 -----TGAACCTTCGCCAGGT 278
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Db 11067 AlaGlyArgArgArgGlyHisGlyAlaMetalaProProArgProGlyThrAlaProGly 11086
 Qy 277 CGATGTAATCAGCCATCATATCATCTCCAGTCCAGCCAGGTCACAGGCCAGCA 218
 Db 11087 ArgAlaArgArgArgTIPAlaArg-----ProGlyArgProArg 11099
 Qy 217 TGAGGTCGAGCGCTCGG-----CCAGCGGGGCAAGAGCGCGCTGGGGCCCGT 164
 Db 11100 ThrArgArgArgAlaAlaProValProArgGlyAlaAlaGlyProValProGlyProSer 11119
 Qy 163 AGAAGCTCTCCA-----GGCGAGCAGCGTGAT-----CATCGCCCT----- 128
 Db 11120 ProAlaSerProSerCysAlaArgGlyArgGlyArgAspAlaHisArgProArgArgGly 11139
 Qy 127 ---TCTTGCTGCCGAACCATATATGGCGCAGCAACGGTCAGGT 89
 Db 11140 SerSerSerArgArgSerAlaGlyProArgArgSerGly 11153

RESULT 11

US-10-084-846A-4
 ; Sequence 4, Application US/10084846A
 ; Publication No. US2004006026A1
 ; GENERAL INFORMATION:
 ; APPLICANT: WEITNAUER, GABRIELE
 ; APPLICANT: MUHLERWEG, AGNES
 ; APPLICANT: TREFFZER, AXEL
 ; APPLICANT: BECHTHOLD, ANDREAS
 ; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
 ; FILE REFERENCE: 1974-005
 ; CURRENT APPLICATION NUMBER: US/10/084, 846A
 ; CURRENT FILING DATE: 2003-02-25
 ; PRIOR APPLICATION NUMBER: PCT/EP01/09815
 ; PRIOR FILING DATE: 2001-08-24
 ; PRIOR APPLICATION NUMBER: DE 101 09 166.4
 ; PRIOR FILING DATE: 2001-02-25
 ; NUMBER OF SEQ ID NOS: 120
 ; SOFTWARE: PatentIn ver. 3.2
 ; SEQ ID NO 4
 ; LENGTH: 19725
 ; TYPE: PRT
 ; ORGANISM: Streptomyces viridochromogenes
 ; FEATURE:
 ; OTHER INFORMATION: Protein 2: amino acid sequence encoded by coding strand 1.
 ; OTHER INFORMATION: Start codon: gat, Start position: nucleotide 2.
 US-10-084-846A-4

Alignment Scores:
 Pred. No.: 2,77e-17 Length: 19725
 Score: 424.00 Matches: 222
 Percent Similarity: 33.2% Conservative: 41
 Best Local Similarity: 28.0% Mismatches: 251
 Query Match: 11.7% Indels: 278
 DB: 4 Gaps: 49

US-10-620-914-44 (1-1947) x US-10-084-846A-4 (1-19725)

Qy 1939 CCTTCTTGGCGCCCTCCGGCGGCCATGTAGAAGGAGCTGTACATGTGACGGGTCCA 1880
 Db 17234 ProSer---ArgProTyrGlyArgProTTPAsnArgSerMetSerAsnTTPProGly--- 17251
 Qy 1879 TGTAGCCCTGAGTGGCGCGG-----GGATGCAGCGCAGCGTCGAAGCCCGCCT 1832
 Db 17252 -----AlaSerTTPArgSerSerArgTTPProGlyCys-----ArgThrSerProArg 17267
 Qy 1831 TCTGGATCAGCT-----CGCGGTAGCGCGGCTGAGGAGCGGAGC----- 1790
 Db 17268 AlaGlyCysAlaTTPGlyArgTTPArgArgAlaGlySerGlyArgSerAlaArgCysSer 17287
 Qy 1789 -----GCCAGATGACGATGCGCGCGCAACCTCTCTTGGCCAGGCACTCGGCCCA--- 1739
 Db 17288 ThrProAlaArgThrGlyCys-----AlaSerSerAlaGlySerGlySerArgProPro 17305

Qy 1738 ---GCTCGTTGG-----CCAGCGGCATAT-----CCAGCCAGTCCA--- 1706
 Db 17306 ThrArgArgTTPProProSerGlySerProArgLeuTyrGlyThrAlaSerProTyr 17325
 Qy 1705 ---CGTGTCTCCATCAGATCAGCT---TCG-----TGTAGTGC 1673
 Db 17326 ThrSerThrTTPThrGlyProSerProGlyProProHisTTPSerSerArgCysThrCys 17345
 Qy 1672 GCGCTTTGAGCTCTCCATGAAGAAGTTGGTGGAGACGGTCAGGTGTTCACCCAGCCAC 1613
 Db 17346 -----TTPTPArgArgAla-----ProArgHis 17353
 Qy 1612 TCTTAGGGTGGCGAAGCCCG-----CCTCGCGCA 1583
 Db 17354 AlaGlyProSerArgArgAlaGlyArgArgSerSerAlaArgCysTTPProArgArg 17373
 Qy 1582 GGTAGTGGGCGAGTTGTGCGCAGGAAGTTCGCGGTGAGGAGTGTGTAGTAGA----- 1529
 Db 17374 GlyArg-----ArgGlnAlaGlyCys-----AlaCysCysSerArgGlyGly 17387
 Qy 1528 -----AGT 1526
 Db 17388 ProHisGlyAspGlyArgTTPArgProTTPArgArgSerAlaHisProSerArgProThr 17407
 Qy 1525 AGTTCTGCTTGGCGCAGCGAGTTCTCCGCCACCGCGTCCATGGTGGCGCGATGTAGT 1466
 Db 17408 ThrThrValCysHisArgCysSerProArgProArgSerThrCysAlaAlaArgAla 17427
 Qy 1465 TCTCAATGG-----GGATGCGCTCGCTTGCATCAGCGCGTACTGCT 1424
 Db 17428 ProArgTTPProProTTPSerThrSerSerCysAlaProPro-----SerThrThrAla 17445
 Qy 1423 TGCCCGCGCAGCGCCCGCCGAAACACACAGCAGCGGCTTGTTCAGAGCAGCACCA--- 1373
 Db 17446 CysSerProSerSerArg-----AlaAlaProProThrProArgArgProArgGlySer 17463
 Qy 1372 -----GGCTCAGAACTTGACGACAGCAGCAGCAGCAGCAGCG 1340
 Db 17464 CysArgThrArgTTPProSerGlySerGlyProArgArgSerThrThrArgThrAlaGly 17483
 Qy 1339 GCTTGGCGCGCTTCTTCCAGAGTGGATGAGCATGTTGCTCTCCACAGCGCGCTGCT 1280
 Db 17484 SerHisCysAlaAlaThrArgProSer---AlaProAlaSerProSerProSerAlaGly 17502
 Qy 1279 CTTCAATTGTGGCGCGCTTGGCGAGCGCTTGCAGGTCTTCCCGCAGTCCACAGCAGCG 1220
 Db 17503 ArgSerSerAlaThrArgTTPGlySerAlaArgArgSerArgProSerProHisSerArg 17522
 Qy 1219 -----CCAGGC 1214
 Db 17523 ThrProProArgGlyArgAlaThrSerTTPSerValArgProAlaCysSerThrGly 17542
 Qy 1213 ACTCAGCAGCAGCAGCAGCTTGCCTATCGCGCCCTGGTAGTACAGCGCGTGTGGAGT 1154
 Db 17543 ProGlyArgProArgSerAlaValCysCysAlaHisArgCysThrAlaProThrGlyArg 17562
 Qy 1153 ACCAGAGCGCTTGGACCAAGTGTGGCTTTCGACAGGAAGCGCGCCAGCTTCT 1094
 Db 17563 ThrArgSer-----ProThrGlyArgGlyGlyAlaAla 17573
 Qy 1093 TCTGTACAGCTCTCAATCGCGGGGTGCAGCGCCCTCGCGAAGAGTGTGCC----- 1043
 Db 17574 SerArgSerProSerThrArgCysGlyAlaSerArgProGlyAlaAlaArgTTP 17593
 Qy 1042 -----ACAGCTCTCAAACT-----CCAGTGTGTAATGGCA 1010
 Db 17594 AlaCysSerSerTTPThrArgProThrThrArgTTPProGlyProAlaAlaProTTPPro 17613
 Qy 1009 CTTCTTCCAGCTCCAGAAAGCGCGACTCGCGGGTTCAGTCCAGCAGCAGCAGCAGTGGC 950
 Db 17614 SerProSerGly-----ArgSerThrAlaThrAlaCysCysSerProValArgProTTP 17631
 Qy 949 CGGCCCGCTGCACACAGCAGGTTTCAGGCGCATTTGCAGCGCGCGCTAGTCAGGCTCAGCAGCG 890

Db	17632	ArgThrAlaSerArgSerAlaAlaTrpCys	::: ::: :::: ::::	GlySerSerSer	17646
QY	889	TGTCCTTGGGTGTATCTCCATCATCGGCTCGGGTGCTCTCCACAGCTGC		CG	830
Db	17647	ProSerTrp		ArgArgProSerThrSer	17655
QY	829	TGTACAGGAAGTGGGGCGGAACATGGCGCGCTCCACCGCTCTCTGTCGAGGCGT		CG	770
Db	17656	---ThrThrAlaTrp		ArgAlaProArg	17663
QY	769	GGCCAACGCTGGGCGAGCGGCCAATCCACGTACTAGGGGGCGC		GCAGCC	719
Db	17664	-----ArgSerAlaArgArgSerHisArgSerIleCysAlaAla		CysAlaAla	17676
QY	718	ACGGCAAGTAGGGATCGAACCTGGGTGTTCTGCCACACCGCTCCAGCTTCGT		CG	659
Db	17677	ThrSerArgThrSerSerProAsnSerArgArgSerThrArgThrSerGlyArgSer		ArgSer	17696
QY	658	CCAGGTAGGCGGCGCTCGGGCGCATGTCAATTGTTGCGATGCGAAGATCGATCGCC		CG	599
Db	17697	ProAlaArgArg		Thr	17701
QY	598	AGAAGAAACGGCGCACCGAGGCATCTGGCGCAGGGCAGGTGCTGTCGCTCAGT		CG	539
Db	17702	ArgArgProThrAlaArgProCysAlaProAlaThrSerTrpArg		CysAlaGlyArg	17720
QY	538	AGAAGTCGGCAGCCACCACCGCCGCTTGGGACAGGTACGAGCAACGCTGTCGATGA		CG	479
Db	17721	ArgThrArgValArgSerGlyProArgGlyCysThrGlyCysAlaArg		CG	17736
QY	478	CGTTGTGGAACGCTGGAATCATCTGTCGAGCGAGTAGGAGAAGGTGATGAGCGTGC		CG	419
Db	17736	-----		CG	17736
QY	418	CCTCAGGGGCGCAAATTTGGCAAGCGTCGGCTCCACGACCTTGGACATCTTCCAGCCCT		CG	359
Db	17737	-----SerLeuArgArgProProArgThrGlyArgSer		ArgSer	17747
QY	358	TGGCCTTCGGCTTCCT		CG	323
Db	17748	TrpCysSerProArgSerGlyThrTyrtPrpGlnTrpTyrAlaArgSerArgThrArg		Arg	17767
QY	322	AGTGGCACAGTCGACCACT--AGATGACATTTGAATTCGCCAGGTCTGTAATCAG		CG	266
Db	17768	ProGlyGlnGlyArgLysGlnAlaArgThrSerArgGlyProLysArg		GlyGlu	17786
QY	265	CCATCATATCGACATTCCTCCCGACGCCACCACCGAGTCAACCCAGGTTCGAGC		CG	206
Db	17787	ProGlnArgMetArgGlyAlaArg--ArgTrpLysGlyGlnProArgThrArgArg		Arg	17805
QY	205	GCTCGG-----CGAGCGGGCGAAMGGCAG-----CGG		CG	176
Db	17806	AlaArgArgSerProProtyrGlyThrHisArgArgArgSerGlyArgGlyArg		GlyArg	17825
QY	175	CCTGGGCCCCGTAGAACCTCTCCAGCGGACGCGTATCGCCCTCTTGTCGCCGA		CG	116
Db	17826	AlaGlyCysArgPepArgGlyProGlyValCysSerGlyArgSerArgAlaAlaCysArg		CysArg	17845
QY	115	-----ACCACATATGGCGCAGAACCGGTCTGAGTCACTCTTCATGC		CG	77
Db	17846	ProProAlaAspSerSerSerThrThrSerProProArgProValThrProtyrCys		Cys	17865
QY	76	TGCTGAGCTTGCAGTCTCCAGGAGAGTTCTTCTTGGTGTAGC-----		CG	32
Db	17866	-----SerArgArgSerGlnAlaGluTrpAlaSerThrCysArgProPro		CG	17881
QY	31	-----TCGACGGCGGGCGCTCAGCACCCGACCCCA		CG	2
Db	17882	ProTrpSerSerAlaSerArgSerSerSerArgPro		CG	17893

RESULT 12
US-10-084-846A-4

```

; Sequence 4, Application US/10084846A
; Publication No. US2004006026A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUHLENWEG, AGNES
; APPLICANT: TREPZER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084,846A
; CURRENT FILING DATE: 2003-02-25
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 4
; LENGTH: 19725
; TYPE: PRT
; ORGANISM: Streptomyces viridochromogenes
; FEATURE:
; OTHER INFORMATION: Protein 2: amino acid sequence encoded by coding strand 1.
; OTHER INFORMATION: Start codon: gat, Start position: nucleotide 2.
US-10-084-846A-4

Alignment Scores:
Pred. No.: 3.07e-16 Length: 19725
Score: 407.00 Matches: 236
Percent Similarity: 34.8% Conservative: 48
Best Local Similarity: 29.0% Mismatches: 237
Query Match: 11.2% Indels: 294
DB: 4 Gaps: 53

US-10-620-914-44 (1-1947) x US-10-084-846A-4 (1-19725)
QY 23 GGCCTGCGAGCTACACCAAGAAGACTTCTCCCTCG-----AGAACCTCAAGC 70
Db 17261 GlyCysAlaThrSerProArgAlaGlyCysAlaTrpGlyArgTrpArgArgAlaGlySer 17280
QY 71 TCAGCAGCATGAGGATGACCTGACCGTCTCGCGCATATGGTGTGCGCAGCAAGAAGG 130
Db 17281 GlyArgSer-----AlaArgCysSerThrProAlaArg--- 17291
QY 131 GCGATGATCAGCTGCTCGCTCGGAGAGCTTCTACGGGCCCCAGGCGCTGCTTGTCTG 190
Db 17292 -----ThrGlyCysAlaSerSerAlaGlySerGlySerArgPro----- 17304
QY 191 CCGCGC-----TGCCCGAGCGCTCGAACCTCATCTGGGTGTGACCTGGGTGGGCACTG 244
Db 17305 ProThrArgArgTrpProProSerGlySerProArgLeu-----Tyr 17319
QY 245 GGGAGAATGTCATATGATGGCTGATTACATCGACTGGCGGAAGTTCAAAGTCCATCTACG 304
Db 17320 GlyThrAlaSer-----ProTyrThrSerThrTrpThrGlyProSerProGlyPro 17336
QY 305 -----TGGTCGACCTGTGCCACTCGCTCGAGGTGGCCAAAGAAGA----- 346
Db 17337 ProHisTrpSer-----SerArgCysThrCysTrpTrpArgArgAlaProArgHis 17353
QY 347 -----AGCGAAAG-----CCNAGGGCT 364
Db 17354 AlaGlyProSerArgArgAlaGlyArgArgArgSerSerAlaArgCysTrpProArgArg 17373
QY 365 GGAAGA-----ATGTCAGGTG----- 382
Db 17374 GlyArgArgGlnAlaGlyCysAlaCysCysSerArgGlyGlyProHisGlyAspGlyArg 17393
QY 383 TGGAGGCCGACGCTTGCCCAATTGGCCGCCCTCAGGGCACCG-----CGACGC 430
Db 17394 TrpArgProTrpArgArgSerAlaHisProSerArgProThrThrThrValCysHisArg 17413
QY 431 TCATCACCTTCTCTACTCGCTCAGCATGATCCACGGTTCACCAACGTCATCGACCCAGG 490

```



```

; SEQ ID NO 3
; LENGTH: 19695
; TYPE: PRT
; ORGANISM: Streptomyces viridochromogenes
; FEATURE:
; OTHER INFORMATION: Protein 1: amino acid sequence encoded by coding strand 1.
; OTHER INFORMATION: Start codon: gga, Start position: nucleotide 1.
US-10-084-846A-3

Alignment Scores:
Pred. No.: 8.56e-15 Length: 19695
Score: 383.50 Matches: 218
Percent Similarity: 34.2% Conservative: 57
Best Local Similarity: 27.1% Mismatches: 268
Query Match: 10.6% Indels: 260
DB: 4 Gaps: 48

US-10-620-914-44 (1-1947) x US-10-084-846A-3 (1-19695)
Qy 1945 AGTTGTCTCT--TCTTGGCGC-----CCTTCGGGGGCGCATGTAGA 1907
Db 4930 SerCysValIleGlyTrpArgHisArgAlaAlaProProAlaCysLeuProCysAla 4949
Qy 1906 -----AGGAGCTGTACATGTGACGGGTCCATGT-----AGCCCTGAGCTGG 1865
Db 4950 CysCysArgCysCysArgCysTrpProArgProCysCysGlyThrAlaProArgAla 4969
Qy 1864 CGCGCGGATGACAGCGCATGTCGAAGCCGCTTTCGATCAGCTCGCGGTAGGCGGCG 1805
Db 4970 Arg-----ThrThrArgArgAspProAlaProGlyValProArgArg-----Gly 4984
Qy 1804 TGAGGGAGCGGAGCGGAGATGACGATGCGCGCGCGGCGCAACCTGTGTCGCGGCACT 1745
Db 4985 MetArgArgMetSerArgArgGlyCysCysSerProSerProTrpTrp----- 5001
Qy 1744 CGGCCAGCTGTTGCGCCAGGGCATATCCAGCCAGTCCACGTGCTCCATCAGAA----- 1691
Db 5002 -----TrpPro-----SerArgGlyProAlaGluArgTrp 5011
Qy 1690 -----TCACCTTGGTGTAGTGCCTTTGAGTCTCTCCATGAAGAAGTTGTGAGA 1637
Db 5012 SerAlaGlyThrTrpAlaSer-----HisAlaTrpTrpAla 5023
Qy 1636 CGTTCAGTTGTCACACAGCCACTTCTGAGGTGGCGAAGCGCGCTCGCGCAGGTAGG 1577
Db 5024 ArgSer-----ProGlySer 5028
Qy 1576 TGGGCACTTGTGCGCAGGAACCTTTCGGGTGAGGACGTTGTAGTAGAAGTAGT----- 1523
Db 5029 PheSerAlaLeuArg-----CysSerAlaSerSerLeuPro 5040
Qy 1522 TCTGCTTCGCGCATGTCGAGTTCCTCCGCCAGCGGTCCA----- 1484
Db 5041 GlySerThrThrProSerProProProSerThrProThrSerIleSerSerProArg 5060
Qy 1483 -----TGTGCGCGCGATGATGTTCTCAATGGGATCCCGTCCGCTTGATCAGCGGT 1430
Db 5061 SerGlyTrpArgCysSerCysSerSerAlaTrpSerSerValThrProThrMetLysAla 5080
Qy 1429 ACTGCTTGCCTCCGCGCAGCGCGCGCGAACCACGACGCGCTTGTGAAGAGCACCGGC 1370
Db 5081 ArgAlaAlaProGlyProProTrpTrpGluSerSerAlaSerAlaCys--ArgSerProSer 5099
Qy 1369 TCACGAACCTTGACGAACA---GCCACACAGCGGCTTGGGCGGCTTCTTACGAAGTGA 1313
Db 5100 AlaValLeuTrpAspThrArgCysThrProAlaThrProArgThrAlaSerAlaSer--- 5118
Qy 1312 TGAGCATGTTGTCTCCACACAGCGGCTGCTCT-----CCATTGGGGGG 1265
Db 5119 -----CysArgSerProSerSerTrpAlaSerProAlaProProSerProCysTrpArg 5136
Qy 1264 CGTTGGCGAGGGCTTGACGGTCTTGGCCAGTCTCCAGCACACCGCGGCGGCACTGCAGCA 1205

```

```

Db 5137 GlyCysTrpAsnAlaAlaGlyCysSerSerProGlyArgAlaProThrProSerTrpGlyPro 5156
Qy 1204 CCCAGCACAGCTTGGCCATCCCGCCCTGTTGTTAGTACAGGCGCTGCTGGAAGTACCAGAGGC 1145
Db 5157 ProProProIleSerProAla-----GlyCysSerArgValSerSerProCysCys 5174
Qy 1144 GCTTGACCAAGAGTTGTTGCTGTTGTCACAGGAAGCGCGCAGCTTCTTCTTCGTACA 1085
Db 5175 AlaAlaValArgProTrpAlaCysCysAlaProSerArgSerProArgSerSerAla 5194
Qy 1084 GCTCTCAATGCGCGGTGACGCCCT----- 1058
Db 5195 SerTrpTrpCysSerSerAlaProProCysAspAlaSerTrpSerGlyArgSerAlaGly 5214
Qy 1057 CGCGAACAGCTGCCACACGCTCTCAAACTCCAGCTGCT-----GAATGGCCACCT 1007
Db 5215 CysProThrAlaAlaCysSerHisSerProAlaSerCysCysProArgSerProPro 5234
Qy 1006 TCTTCAGTCCAGAACGCGCGACTGCGCGGTTGTCAGTCCACCGCACCA----- 956
Db 5235 SerSerAlaSerIleSerSerSerAlaProSerCysSerAlaProSerAlaArgArgPro 5254
Qy 955 -----CCTGCGCGCGCCCTCTGCACCA-----GCAGTTCCAGGCATGCGAGC 914
Db 5255 ArgArgProTrpArgThrProAlaAlaAsnCysArgAsnSerSerProProTrpCysCys 5274
Qy 913 CGCGCTAGTCA-----GGGTCCAGCA 893
Db 5275 ArgArgSerSerProArgSerAlaArgProThrCysCysAsnSerAlaGlyAlaAla 5294
Qy 892 CCGTGCTCTTGGGTTGATCTCCATCATCTCCATATCCGGTCCGGGTCTCCACAGCT 833
Db 5295 ProCysGlyCysGly-----ProAlaSer 5302
Qy 832 GCGTGTACAGAAAGTGGCGGGAACATGGCGCGCGCTCCACCGGCTCTCTGTCGAGGG 773
Db 5303 ProCysTrpTrpProSerSerAlaAsnTrpArgAlaAlaArgArgGlnArgCys--- 5321
Qy 772 CGTGGCCACGCTGGCGAGCGGCCAATCCACAGTAGTAGGGGCGCGGCGCACCGCA 713
Db 5322 -----ArgTrpSerAla-----SerThrProCysAlaSerAlaCysSerThrAla 5336
Qy 712 -----CGTAGGGATCGAACCCCTGGGTGTTCT 686
Db 5337 AlaAspProAsnTrpSerSerProSerGlySerAsnSerAlaSerHisProArgCysSer 5356
Qy 685 ---GCTCCACACCGCTCCAGCTTCTGCTCCAGTAGGCGGCGCTCGGGCGCGAGT 629
Db 5357 ProCysTrpSerSerProCysAlaProProSerProArg-----ArgCys 5372
Qy 628 CAATGTTGT-----CGATGTCCAGATCGATCGCGCAGAGAAACCGCGCG 584
Db 5373 SerThrCysSerThrAlaProArgArgAlaProArgProProValGlnArgValPro 5392
Qy 583 ACCAGGCGATCTGGCGGCGGCGAGTGTGCTGCTCCAGTAGGCGGCTCGGCAACGC 524
Db 5393 SerPheAlaAspAsp---GlyAlaAlaAsnSerAlaArgGluMetProGluArgVal 5411
Qy 523 CCACGAGCGCTCTGGGACA----- 503
Db 5412 IleProSerArgIleProThrLysArgLeuSerSerLysGluCysArgSerValArgAla 5431
Qy 502 GGTACGACAGCCCTGGT-----CGATGACGTTGTGGAACGGTG 464
Db 5432 GlySerSerLysProGluAsnArgGluValGlyTrpCysArgProThrCysThrSerVal 5451
Qy 463 GAATCATCTGTCGAGGAGTAGGAGAGTGTAGCGTCCGCGTCCCTCAGGGGGCGCAA 404
Db 5452 LeuSerAlaSerSerTrpArgVal-----AlaSerAsnGlyProTrpAsnAla--- 5468
Qy 403 ATTGCAAGCTGCGCTCCAGCACCTTCGACATTTCTTCAGCCCTTGGCTTCGCTTCT 344
Db 5469 ---AlaSerIleArgProLysArgProArgHisThrSerSerAlaGlySerProProAla 5487

```

QY 343 TCT-----TGCCCA----- 335
 Db ThrSerProLeuArgLysTrpProCysGlyArgArgSerArgArgGlyGlyAlaAasp 5507
 QY 334 -----CCTCCGACAGCGAGTGCCACA--- 314
 Db ArgArgAasnSerIleCysCysMetArgAlaProGlyThrArgAlaArgThrAla 5527
 QY 313 GGTGACACAGTAGA-----TGGAAGTGAAGT---TCGCCA----- 281
 Db GlyArgProArgHisThrCysSerGlyThrTrpTrpAlaGlyThrCysSerProTrpArg 5547
 QY 280 -----CGTCGATGTAATCAGCATCATCATATCGACATCT 248
 Db SerGlyArgAlaAlaThrAlaCysSerAlaAlaArgTrpAasnSerProSer--AlaThrCysA 5567
 QY 247 CCCAGTG-----CCACACCCAGGTCAACCCAGATGAGTTCG 209
 Db rGProIleArgThrAlaProAlaArgSerProProThrThrThrGluArgArg-Ser 5586
 QY 208 AGCGCTCGCCGAGCGGG-----CAGCAAGGCGACGCGCGCTGGG 170
 Db ThrGly-----GlyGlyTrpAlaProAlaSerSerAlaAlaMetArgSerArgProGly 5604
 QY 169 GCCCGTAGA-----AGCTTCCAGGCGAGCAGCGTGATCATATCCCTTCTTGTCTGCCGA 116
 Db AlaAlaSerGlyProAlaSerProGlySerAlaArg---TrpProProArgGlyCysArg 5623
 QY 115 ACCACATATGGCGAGACGGTCAGGTCACTCTTCATGCTGCTGAGCTTGAGCTTCTCCA 56
 Db ArgSer-----SerArgCysThrAlaValThrSerArgCysSerArgAlaSerPro 5641
 QY 55 GGG 53
 Db 5642 Gly 5642

RESULT 14
 US-10-118-495-3
 ; Sequence 3, Application US/10118495
 ; Publication No. US20030074688A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Benning, Christoph
 ; APPLICANT: Riekhof, Wayne
 ; TITLE OF INVENTION: Compositions and Methods for the Production of Betaine Lipids
 ; FILE REFERENCE: MSU-06897
 ; CURRENT APPLICATION NUMBER: US/10/118,495
 ; CURRENT FILING DATE: 2002-04-08
 ; PRIOR APPLICATION NUMBER: 60/283,812
 ; PRIOR FILING DATE: 2001-04-13
 ; NUMBER OF SEQ ID NOS: 42
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 3
 ; LENGTH: 416
 ; TYPE: PRT
 ; ORGANISM: Rhodobacter sphaeroides
 US-10-118-495-3

Alignment Scores:
 Pred. No.: 6.82e-15 Length: 416
 Score: 379.50 Matches: 127
 Percent Similarity: 43.1% Conservative: 68
 Best Local Similarity: 28.1% Mismatches: 182
 Query Match: 10.5% Indels: 75
 DB: 4 Gaps: 20

US-10-620-914-44 (1-1947) x US-10-118-495-3 (1-416)

QY 691 ACCCAGGTTTCGATCCCTAGTCGCGTGGCTGGCGGCCCTACTACTAGTGTGATGGC 750
 Db 2 ThrGlnPheAlaLeuThrHisLeuPro-----AlaProPro-----Vala 15

QY 751 GCGTGGCCAGCGTGTGGCCACGCGCTGCAC-----GAGAGCGC 789
 Db Arg-----GlnIleGlyAlaAlaValHisArgThrSerLeuLeuSerAlaGlyLeu 33
 QY 790 GTGAGCGCGCGCCATGTTCCCGCCACCTTC-----CTGTACACGAGTCGTGG 840
 Db MetGluArg-----MetPheSerArgLeuPheHisGlyLeuValTrpProGlnIleTrp 51
 QY 841 GAGGACCCCGAGCGGATATGGAGGTGATGAGATCAACCCCAAGGACACCGTGTGACC 900
 Db GluAaspProAlaValAaspMetAlaAlaAlaIleArgProGlyAaspArgLeuVala 71
 QY 901 CTGACTAGCGCGCTCAATGCCCTGAACCTGTCTGTGTCGAGGGCGCGCCAGGTGTG 960
 Db IleAlaSerGlyGlyCysAasnValLeuSerTrpLeuThrGlnGlyProGlySerIleLeu 91
 QY 961 TCGGTGACTGCACACCCCGCGCAGTCGCGCTTCTGGAGCTGAAGAAGTGGCGCATTCAG 1020
 Db AlaValAaspLeuSerProAlaHisValAlaLeuGlyArgLeuLeuAlaAlaArg 111
 QY 1021 CAGCTG---GAGTTTGGAGCAGTGTGGCAGCTGTTCCGCGAGGCGGTGCACCCGCGCAT 1077
 Db ThrLeuProAaspHisAlaAlaPhePheAaspLeuPheGlyArgAlaAaspLeuProGlyAasn 131
 QY 1078 GAGGAGCTGTACGAGAAGAGTGGCGCCCTTCTGTCTGCGCAACACGACCACTTCTGG 1137
 Db AlaAlaLeuTrpAaspArgHisIleAlaProAlaAlaAaspGlyArgSerArgTrp 151
 QY 1138 -----TCCAGCGCTCTGGTACTTCCAGCACGCGCTGTACTACCAG 1179
 Db GluAlaArgSerProPheGlyArgArgIleGlnLeuPheGluArgGlyPheTrpArgHis 171
 QY 1180 GCGCGCATGGCAAGCTGTGCTGGTGTCTGAGTGCCTGCGCGTGGTGTGGACTGGCG 1239
 Db GlyAlaLeuGlyArgPheIleGlyAlaAlaHisThrLeuAla-----ArgAlaAaspGly 189
 QY 1240 AAGACCTCAAGCGCCTCGCCCAACGCGCCCAATGAGGAGGACGCGCGTCTGTGGAC 1299
 Db ThrAaspLeuArgGlyPheLeuAaspCysProAaspIleGluAlaGlnArgSerPheTrp 209
 QY 1300 AGCAACATGCTCCTCCACTTCGTGAAGAAGCGGCGCCCAAGCGCGTGTGTGCTGTCTGTC 1359
 Db AlaHisIle-----GlyPro-----LeuPheGlu 217
 QY 1360 AAGTTCGTGAGCTGTGCTCTTCAACAAGCGCGTGTGTGTTTCGCGCGCGGTGCGG 1419
 Db AlaProValValGlnAlaLeuAlaArgProAlaAlaLeuPheGlyLeuGlyLeuPro 237
 QY 1420 GGCAGCAGTACGCGCTGATCAAGCGCGACGCG-----ATCCCATTTGAGAAC 1467
 Db ProAlaGlnTrpAlaLeuLeuAlaGlyAaspGlyAaspGlyAaspValLeuProVal----- 255
 QY 1468 TACATCGCGCGCACCATGGACGCGGTGGCGAGAACTCGCAC----- 1509
 Db LeuArgGlnArgLeuHisArgLeuLeuCysAaspPhe 267
 QY 1510 GTGCGCAAGCAGAACTACTTCTACTCAACTGCTCACCAGCAAGTTCCTGCGCGACAAC 1569
 Db ProLeuArgGluAasnTrpPheAlaPheGlnAlaIleAlaArgArgTrpProArgProGly 287
 QY 1570 -----TGCCCCACCTTACCTGCGGAGCGCGCTTCCGCCACCTCCAAAGATGGCGTG 1620
 Db GluGlyAlaLeuProProTrpLeuGluProThrAlaPheGluThrLeuArgGluAasn--- 306
 QY 1621 GTGACAACTGACCGCTCTCCCACTTCTCATGAGGAGGTCAAAGCGCGCACCTAC 1680
 Db AlaGlyArgValGlnIleGluAasnArgSerLeuThrGluAlaLeuAlaGluProGlu 326
 QY 1681 ACCAAGTG-----ATTCTGATGACACGCTGAGCTGGCTGATATGCCGTGGCC 1731
 Db GluSerIleHisGlyPheThrLeuLeuAaspAlaGlnAaspTrpMetThrAaspAlaGlnLeu 346
 QY 1732 AACGAGCTGCGCGAGTGCCTGGCCAAAGCAGAGTGTGGCGCGCGCGCATCTGTCATCTGCGCG 1791

```
Db 347 ThrAlaLeuTrpArgGlnValThrArgThrAlaAlaProGlyAlaArgValIlePheArg 366
Qy 1792 -----TCGCGCTCCCTCAGCCCGCCCTTACCGCGAGCTGATCCAGAGCGGGGCTTC 1842
Db 367 ThrGlyAlaAlaAspLeuPro-----GlyArgValProGluGluIleLeuGly 384
Qy 1843 GACGTGGCTGCATCCCGCGCGCCACTCAGGCG-----TACATGGACCGGCTCAACATG 1896
Db 385 HisTrpArgAlaAspArgAlaAlaGlyGlnAlaGlyHisAlaAlaAspArgSerAlaIle 404
Qy 1897 TACAGCTCTTCTACATGGCCCGCGGAAGGCGCC 1932
Db 405 TyrGlyPheHisLeuTyrArgArgAspAla 416

RESULT 15
US-10-620-914-3
; Sequence 3, Application US/10620914
; Publication No. US20040093639A1
; GENERAL INFORMATION:
; APPLICANT: Benning, Christoph
; APPLICANT: Riekhof, Wayne
; APPLICANT: Klug, Rouven
; TITLE OF INVENTION: Compositions and Methods for the Production of Betaine Lipids
; FILE REFERENCE: MSU-07769
; CURRENT APPLICATION NUMBER: US/10/620,914
; CURRENT FILING DATE: 2003-07-16
; PRIOR APPLICATION NUMBER: 10/118,495
; PRIOR FILING DATE: 2002-04-08
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
; LENGTH: 416
; TYPE: PRT
; ORGANISM: Rhodobacter sphaeroides
US-10-620-914-3

Alignment Scores:
Pred. No.: 6.82e-15 Length: 416
Score: 379.50 Matches: 127
Percent Similarity: 43.1% Conservative: 68
Best Local Similarity: 28.1% Mismatches: 182
Query Match: 10.5% Indels: 75
DB: 4 Gaps: 20

US-10-620-914-44 (1-1947) x US-10-620-914-3 (1-416)
Qy 691 ACCGAGGTTCGATCCCTCAGTGGCGTGGCTGCGCGCCCTACTACGTGTGGATTGGC 750
Db 2 ThrGlnPheAlaLeuThrHisLeuPro-----AlaProPro-----ValAla 15
Qy 751 CGCTGCCAGCGTGGCCAGCCCTGCAC-----GAGGAGCGC 789
Db 16 Arg-----GlnIleGlyAlaAlaValHisArgThrSerLeuLeuSerAlaGluGlyLeu 33
Qy 790 GTGGAGCGCGCCCATGTTCCCGCCCACTTC-----CTGTACAGCAGTCTGG 840
Db 34 MetGluArg-----MetPheSerArgLeuPheHisGlyLeuValTyrProGlnIleTrp 51
Qy 841 GAGGACCCCGAGCGCGATGATGAGGTGATGGAGATCAACCCCAAGGACACCGTGTGACC 900
Db 52 GluAspProAlaValAspMetAlaAlaLeuAlaIleArgProGlyAspArgLeuValAla 71
Qy 901 CTGACTAGCGCGGTGCAATGCCCTGAACTGTGGTGCAGGGGCGCGCGAGTGTG 960
Db 72 IleAlaSerGlyGlyCysAsnValLeuSerTyrLeuThrGlnGlyProGlySerIleLeu 91
Qy 961 TCGGTGGAGTGAACCCCGCGCAGTCGCGGCTTCTGGAGCTGAAGAGTGGCCATTGAG 1020
Db 92 AlaValAspLeuSerProAlaHisValAlaLeuGlyArgLeuLeuLeuAlaAlaArg 111
Qy 1021 CAGCTG---GAGTTTGAGGACGTGTGGCAGCTGTTCGGCGAGGCGGTGCACCCCGCGCAT 1077
```

```
Db 112 ThrLeuProAspHisAlaAlaPhePheAspLeuPheGlyArgAlaAspLeuProGlyAsn 131
Qy 1078 GAGGAGCTGTACGAGAAGAGTGGCGCCCTTCTGTCGCAACACGACCACTCTGG 1137
Db 132 AlaAlaLeuTyrAspArgHisIleAlaProAlaLeuAspGlyArgSerArgArgTyrTrp 151
Qy 1138 -----TCCAAGCGCTCTGGTGTACTTCCAGCAGCGCCCTGTACTACCAG 1179
Db 152 GluAlaArgSerProPheGlyArgArgIleGlnLeuPheGluArgGlyPheTyrArgHis 171
Qy 1180 GCGCGCATGGCAAGTGTCTGGTCTGCTGAGTGTCTGGCGGTGGTGGTGGTGGTGG 1239
Db 172 GlyAlaLeuGlyArgPheIleGlyAlaAlaHisThrLeuAla-----ArgAlaAlaGly 189
Qy 1240 AAGACCGTCAAGCGCTCGCCCAAGCGCCACATGAGGAGGAGGAGCGCGCTGTGGAC 1299
Db 190 ThrAspLeuArgGlyPheLeuAspCysProAspIleGluAlaGlnArgSerPhePheTyr 209
Qy 1300 AGCAACATGCTCATCCACTTCTGTGAAGAACGCGGCCCAAGCGCGTGTGTGGCTTC 1359
Db 210 AlaHisIle-----GlyPro-----LeuPheGlu 217
Qy 1360 AAGTTCGTGAGCTGTGTCTTCAACAGGCCCTGTGTGTGGCGGGCGGTGGCG 1419
Db 218 AlaProValValGlnAlaLeuAlaArgArgProAlaAlaLeuPheGlyLeuGlyIlePro 237
Qy 1420 GGCAGCAGTACGCGGTGATCAAGCGCGGAGCGG-----ATCCCATTTGAGAAC 1467
Db 238 ProAlaGlnTyrAlaLeuLeuAlaGlyAspGlyAspGlyAspValLeuProVal----- 255
Qy 1468 TACATCGCGCGCACCATGAGCGCGTGGCGGAGCAACTCGCAC----- 1509
Db 256 -----LeuArgGlnArgLeuHisArgLeuLeuCysAspPhe 267
Qy 1510 GTGCGCAAGCAGAACTACTTACTACAACTGCTCACCAGCAAGTCTCTCGCGGACAC 1569
Db 268 ProLeuArgGluAsnTyrPheAlaPheGlnAlaIleAlaArgArgTyrProArgProGly 287
Qy 1570 -----TGCCCCCACTACCTCGCGGAGCGCGCTTGGCCACCTCAAGAGTGGCGTG 1620
Db 288 GluGlyAlaLeuProProTyrLeuGluProThrAlaPheGluThrLeuArgGluAsn--- 306
Qy 1621 GTGCAACAACCTGACCGCTCTCCCACTTCTTCATGAGGAGAGCTCAAGCGCGCACCTAC 1680
Db 307 AlaGlyArgValGlnIleGluAsnArgSerLeuThrGluAlaLeuAlaAlaGluProGlu 326
Qy 1681 ACCAAGTGT-----ATTCTGATGACACAGTGGACTGCTGGATATGCCGTGGCC 1731
Db 327 GluSerIleHisGlyPheThrLeuLeuAspAlaGlnAspTrpMetThrAspAlaGlnLeu 346
Qy 1732 AACGAGCTGGCGGAGTGTGCGCAAGCAGGTGGCGCGCGGCGCATCTGTCTCTGGCGC 1791
Db 347 ThrAlaLeuTrpArgGlnValThrArgThrAlaAlaProGlyAlaArgValIlePheArg 366
Qy 1792 -----TCGCGCTCTCTCAGCGCCCTTACGCGAGCTGATCCAGAGCGCGGCTTC 1842
Db 367 ThrGlyGlyAlaAlaAspLeuPro-----GlyArgValProGluLeuLeuGly 384
Qy 1843 GACGTGGCTGCATCCCGCGCGCCACTCAGGCG-----TACATGAGCGCGCTCAACATG 1896
Db 385 HisTrpArgAlaAspArgAlaAlaGlyGlnAlaGlyHisAlaAlaAspArgSerAlaIle 404
Qy 1897 TACAGCTCTTCTACATGGCCCGCGGAAGGCGCC 1932
Db 405 TyrGlyPheHisLeuTyrArgArgAspAla 416
```

Search completed: March 14, 2006, 01:57:05
Job time : 459.5 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: March 14, 2006, 01:58:46 ; Search time 94 Seconds
(without alignments)
2880.355 Million cell updates/sec

Title: US-10-620-914-45

Perfect score: 3463
Sequence: 1 MGSGRDRPASVTKKNFSL.....RVNYSFYARRKGAKKDN 648

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA_Main:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
- 3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
- 4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
- 6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3463	100.0	648	4	US-10-620-914-45
2	891	25.7	752	5	US-10-741-849-7021
3	850.5	24.6	908	4	US-10-620-914-50
4	379.5	11.0	416	4	US-10-118-495-3
5	379.5	11.0	416	4	US-10-620-914-3
6	326.5	9.4	416	4	US-10-118-495-29
7	326.5	9.4	416	4	US-10-620-914-29
8	321	9.3	415	4	US-10-118-495-41
9	321	9.3	415	4	US-10-620-914-41
10	292.5	8.4	416	4	US-10-118-495-33
11	292.5	8.4	416	4	US-10-620-914-33
12	123	3.6	2284	5	US-10-684-141-58
13	123	3.6	2284	5	US-10-810-486-58
14	120	3.5	221	4	US-10-118-495-35
15	120	3.5	221	4	US-10-620-914-35
16	119	3.4	2286	5	US-10-684-141-54
17	119	3.4	2286	5	US-10-810-486-54
18	112.5	3.2	672	4	US-10-282-122A-68178
19	111	3.2	198	5	US-10-617-320-4622
20	110.5	3.2	298	5	US-10-450-763-38334
21	107	3.1	255	4	US-10-220-381-24
22	107	3.1	255	5	US-10-988-984-2
23	107	3.1	691	5	US-10-732-923-9665
24	106.5	3.1	670	4	US-10-282-122A-69771
25	106	3.1	472	3	US-09-974-298-120
26	106	3.1	472	4	US-10-247-671-186
27	106	3.1	472	4	US-10-211-462-201

28	106	3.1	472	4	US-10-468-406-2	Sequence 2, Appli
29	106	3.1	472	5	US-10-370-715B-256	Sequence 256, App
30	106	3.1	852	4	US-10-437-963-108559	Sequence 108559, A
31	106	3.1	2039	4	US-10-369-493-21994	Sequence 21994, A
32	104.5	3.0	663	4	US-10-282-122A-59897	Sequence 59897, A
33	104.5	3.0	664	4	US-10-389-647-469	Sequence 469, App
34	104	3.0	225	4	US-10-118-495-42	Sequence 42, Appl
35	104	3.0	225	4	US-10-620-914-42	Sequence 42, Appl
36	103.5	3.0	1064	5	US-10-732-923-8191	Sequence 8191, Ap
37	103	3.0	908	4	US-10-437-963-132823	Sequence 132823, A
38	101.5	2.9	663	3	US-09-815-242-10070	Sequence 10070, A
39	101.5	2.9	663	4	US-10-287-274-332	Sequence 332, App
40	101.5	2.9	663	4	US-10-282-122A-56452	Sequence 56452, A
41	101	2.9	652	4	US-10-282-122A-73132	Sequence 73132, A
42	101	2.9	679	4	US-10-282-122A-48489	Sequence 48489, A
43	100	2.9	323	4	US-10-437-963-162897	Sequence 162897, A
44	99.5	2.9	471	4	US-10-375-884-1	Sequence 1, Appli
45	99.5	2.9	663	3	US-09-815-242-14080	Sequence 14080, A

ALIGNMENTS

RESULT 1
US-10-620-914-45
; Sequence 45, Application US/10620914
; Publication No. US20040093639A1
; GENERAL INFORMATION:
; APPLICANT: Benning, Christoph
; APPLICANT: Riekhof, Wayne
; APPLICANT: Klug, Rouven
; TITLE OF INVENTION: Compositions and Methods for the Production of Betaine Lipids
; FILE REFERENCE: MSU-07769
; CURRENT APPLICATION NUMBER: US/10/620,914
; CURRENT FILING DATE: 2003-07-16
; PRIOR APPLICATION NUMBER: 10/118,495
; PRIOR FILING DATE: 2002-04-08
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 45
; LENGTH: 648
; TYPE: PRT
; ORGANISM: Chlamydomonas reinhardtii
US-10-620-914-45

Query Match 100.0%; Score 3463; DB 4; Length 648;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 648; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MGSGRDRPAS	YTKKNFSL	EKLSNKKD	LT	TVLRHMWFGSKGDDHAA	RLSFYGPQAA	60
Db	1	MGSGRDRPAS	YTKKNFSL	EKLSNKKD	LT	TVLRHMWFGSKGDDHAA	RLSFYGPQAA	60
QY	61	APAAALAESNL	IWDLGGT	GENVDMADY	IDLAKFKSI	YVVDLCHSLCEVAKKAKAK	120	
Db	61	APAAALAESNL	IWDLGGT	GENVDMADY	IDLAKFKSI	YVVDLCHSLCEVAKKAKAK	120	
QY	121	GWKNQVQV	EADACQAF	PECTATLIT	FSYSLTWI	PPFHNVIDOACSYLSQDGLVGVADPY	180	
Db	121	GWKNQVQV	EADACQAF	PECTATLIT	FSYSLTWI	PPFHNVIDOACSYLSQDGLVGVADPY	180	
QY	181	VSGKYDLP	LQMPWSRR	FFWRSIFD	INDIGERRAYLEQKLERVWEQNTQGS	240		
Db	181	VSGKYDLP	LQMPWSRR	FFWRSIFD	INDIGERRAYLEQKLERVWEQNTQGS	240		
QY	241	LRAPIYVW	IGRLPSVGH	ALHEER	VERPMPPTFLYTQSWEDPEPDMVMEINPKD	TVLT	300	
Db	241	LRAPIYVW	IGRLPSVGH	ALHEER	VERPMPPTFLYTQSWEDPEPDMVMEINPKD	TVLT	300	
QY	301	LTSGGCNAL	NLLVQAG	QVSVDCN	PAQSALLLEKVAIQOLEFEDVWOLF	FGSGVHPRIE	360	
Db	301	LTSGGCNAL	NLLVQAG	QVSVDCN	PAQSALLLEKVAIQOLEFEDVWOLF	FGSGVHPRIE	360	

361 ELYEKKLAPLPSOTSHNFWSKRLWYFOHGLYYOGGKGLCWVLQCLAVLGLKTKVKRLA 420
Db 361 ELYEKKLAPLPSOTSHNFWSKRLWYFOHGLYYOGGKGLCWVLQCLAVLGLKTKVKRLA 420
Qy 421 NAPTMBEQRBLWSNMLIHFWKNGPKPLWLVFKFVSLVLFNKAIVLWFGGVPKQYALI 480
Db 421 NAPTMBEQRBLWSNMLIHFWKNGPKPLWLVFKFVSLVLFNKAIVLWFGGVPKQYALI 480
Qy 481 KADGIPENYIARTMDGVAENSHVRKQNYFYNCCTGKFLRDNCPTVLRFAAATLKSGV 540
Db 481 KADGIPENYIARTMDGVAENSHVRKQNYFYNCCTGKFLRDNCPTVLRFAAATLKSGV 540
Qy 541 VDNLTSTNFMELKARTYTKVILMDHVDWLDMPVANELAECLAKQVAPGGIVWRSAS 600
Db 541 VDNLTSTNFMELKARTYTKVILMDHVDWLDMPVANELAECLAKQVAPGGIVWRSAS 600
Qy 601 LSPPYAELIQKAGFVRCIRRTATQGYMDRVNMYSSFFYMARCKGAKKDN 648
Db 601 LSPPYAELIQKAGFVRCIRRTATQGYMDRVNMYSSFFYMARCKGAKKDN 648

RESULT 2
US-10-741-849-7021
; Sequence 7021, Application US/10741849
; Publication No. US20050019931A1
; GENERAL INFORMATION:
; APPLICANT: Roemer, Terry
; APPLICANT: Jiang, Bo
; APPLICANT: Boone, Charles
; APPLICANT: Bussey, Howard
; TITLE OF INVENTION: Nucleic Acids Encoding Anti-fungal Drug Targets and Methods of
; TITLE OF INVENTION: Use
; FILE REFERENCE: 10182-023-999
; CURRENT APPLICATION NUMBER: US/10741,849
; CURRENT FILING DATE: 2003-12-19
; PRIOR FILING DATE: 2003-07-16
; PRIOR FILING DATE: 2003-07-16
; PRIOR FILING DATE: 2002-12-19
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 7021
; LENGTH: 752
; TYPE: PRT
; ORGANISM: Candida albicans
US-10-741-849-7021

Query Match 25.7%; Score 891; DB 5; Length 752;
Best Local Similarity 32.7%; Pred. No. 4.4e-80;
Matches 229; Conservative 98; Mismatches 232; Indels 142; Gaps 21;
Qy 51 LESFYGPQAAAF-----AARLAERSNLIWVDLGGGTGGENMDMADYI 92
Db 70 LESFYKQAHYNTREFLLKRGQECURLAISHLPKKKDLIWDIGGGTGSNIEFMDIS 129
Qy 93 DLAK-FKSIYVVDLCHSLCEVAKKAKAGKMNQVVEADACOPAPPEGTATLITFSYSL 151
Db 130 KISENFKAVLVLDLPSLCEVAKARFEHEWNVHVLVADACOPTIDYDSADLITFSYSL 189
Qy 152 TMIPPFHNVIDQACSYLSQDLGVADPYVS-----GKYD-----LPLRQMPWSRRFFWR 201
Db 190 SMIPFTFAADNAVSKLDMEGIIATVDGQSSDTSNGRINTVGLVNRDIPILNFRW 249
Qy 202 SIPDIDNIDIGPERAYLEOKLERVWQNTQ---GSIPIYVWMLRAPYVWIGRLPSVGH 257
Db 250 IWFADKVFLLDSRRNLYEYFGTVKSLNSYNKALGKI-----PYIWIWIGCDKSKSH 301
Qy 258 ALHEE-----RVERPPMPFPT-----WIGR-----LPSVGHALH 417
Db 302 TILERLNLATESPYLAPTTTPIANQLEDIPISKGHEAALINLQNLPPYSMYQKEYWR 361
Qy 274 -----FLYQSWEDDPEDMEVMEINPKDVTULTSGGCHALNL--LVQG 315
Db 362 VYDEMNPLEYQKNOYIYFTWEDDPREDHKLNFNTSDDTVAITTSYASLPTP 421

Qy 316 AGQVSVDCNPAQSALLELKVAIQOLEFEDVWQLFEGEGVHPRIEELYEKKLAPFLSOTS 375
Db 422 PKKIHADVLPNCQNHLELKLASFRCLSQEQIWSMEGEGKIEFNFDLLDITLAPHMSSNA 481
Qy 376 HNPWSKR--LWTFQHGGLYYOGGKGLCWVLQCLAVLGLKTKVKRLANAPTMBEQR 430
Db 482 FOYMDKGPKTSGKGLY---DTGFSRWALRLSRYVFKVCGVSKYVEELCAATTMBEQLR 538
Qy 431 LWDNSNMLIHFWKNGPKPLWLVFKFV--SLVLFNKAIVLWFGGVPKQYALIKADGIPEN 489
Db 539 INNEHL-----KPT--LFPVVGSLVGNMFLWALGVPAQAALM---GPSVIK 584
Qy 490 YIARTMDGVAENSHVRKQNYFYNCCTGKFLRDNCPTVLRFAAATLK-----SGVV 541
Db 585 YVVDLTDPILKRSWISNDNYFYLCNMGRYTKNCCPDYLTGKFNLSSTPAATASGSSPI 644
Qy 542 DNLTVSTNFMELKARTYTKVILMDHVDWLD---MPVANELAECLAKQVAPGGIV 595
Db 645 DNLRIHTDTLINEVFGRLKEKSIITAIIMDHMDWFDNGRDAINEIT-ALKRCLAPGGRVL 703
Qy 596 WRSASLSPPYAELIQKAGF-DVRCIRRTATQGYMDRVNMYSS 635
Db 704 LRSASTKPMYLTFTKRLGFOEENVVVRQPGSSIDRVNMYAN 744

RESULT 3
US-10-620-914-50
; Sequence 50, Application US/10620914
; Publication No. US20040093639A1
; GENERAL INFORMATION:
; APPLICANT: Benning, Christoph
; APPLICANT: Riekhof, Wayne
; APPLICANT: Klug, Rouven
; TITLE OF INVENTION: Compositions and Methods for the Production of Betaine Lipids
; FILE REFERENCE: MSU-07769
; CURRENT APPLICATION NUMBER: US/10620,914
; CURRENT FILING DATE: 2003-07-16
; PRIOR FILING DATE: 10/118,495
; PRIOR FILING DATE: 2002-04-08
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 50
; LENGTH: 908
; TYPE: PRT
; ORGANISM: Neurospora crassa
US-10-620-914-50

Query Match 24.6%; Score 850.5; DB 4; Length 908;
Best Local Similarity 30.4%; Pred. No. 7.3e-76;
Matches 222; Conservative 105; Mismatches 235; Indels 169; Gaps 18;
Qy 69 RSNLIWVDLGGGTGGENMDMADYIDLAK-FKSIYVVDLCHSLCEVAKKAKAGKMNQV 127
Db 178 RRPXIWDVGGGTGGMNIEAMAKFVNVEFFKTVLYVDFSPSLCEVAKKARFARLWENVRV 237
Qy 128 VEADACQFA-----PPEGT-----ATLITFSYSLTILPFP 157
Db 238 ICTDARKFLEDYEDVDEGESGSDSPSLSGHWGETKPGRHAGAEILITMSYLSLWMPDY 297
Qy 158 HNVIDQACSYLSQDLGVADPYVSQKYD-----LPLRQMPWSRRFFWRSFIFDIN 208
Db 298 FSIIDSLSLESLAPHGLIAVVDVFAQSKVDPTFRNTYTGGLNMRHVGVFARNFWSWFDADR 357
Qy 209 IDIGPERRAYLEOKLERVW---EQNTQGISPIYVW---LRAPYV----- 247
Db 358 VSLEPARRDYLYRFGTTLTVNARNLTGAIPYIWIWIGLUKPKFSTSSLPHEIHEIDAI 417
Qy 248 -----WIGR-----LPSVGHALH 460
Db 418 ATSPRSSPLVKGHSSATNALAPAVGRTPAEMRSKAFNTAIENISANLPLSFYQNH 477
Qy 261 -----EERVERPPMPFPTFLYQSWEDDPEDMEVMEINPKDVTULTSGGCHALNLVQ 314

Db 478 HWRIYDDQLPKTQFNDEYIYAFTWEDSRVRELLNLGPDVVVLAITSAGDNILSYMQ 537
QY 315 GAGQVSVDCNPAQSALLEKKVAIOLEPEDVWQFGEVHPRIEBELYEKKLAPFLSQT 374
Db 538 SPARVAIDLNPAQNHLLEKVASFTTLDYFDVWKFGEKHPDFRSLILSKLSPHLSGR 597
QY 375 SHNFWKRLWYFO-----HGLYOGGKGLCWVLOCLAVLGLGKTVKRLANAPTMEQRR 430
Db 598 AFQYWLNSAHIFTDPAGRGVLDYGGSRYAIRFRWISTLFFCRSAVRRLLSTPTEGQRS 657
QY 431 LWDNMLIHVKNVGNPLVWLFVKFYS-LVLFNKAVLWFGGVPKGQYALIKAD-----483
Db 658 IYHTKI-----RPC-LIMRPNGLVSSDAFLMSALGPKVKNQVAMIEDIYHRSI 706
QY 484 -----GPIENIYARTMDGVAENSHVRKQNYFYNNCLTGKFLRNCPTYL 529
Db 707 SSTTSSKEKPSRAEAILHYTTSTLDPVLSHLSNDNPYLVCLVQVYTRQCHPYLS 766
QY 530 EAPATLKS-GVVDNLTVSTNFMEL---KARTYKVIIMDHVDWLDMP-----575
Db 767 PAAHSILSAPGAFDGLRIHTDEIQEVLARFQGTGLTAVVMDGMDWDFPSPPEEKEGRG 826
QY 576 VANELAECLAKOVA PGGIVWRSASLSPPYAELIQKAGDVRCI-----RRATQGYMDRV 630
Db 827 KAREQVRLNRALKVGGKVLRSAGVEPTVRFVBEGFGARRVGRSGRQDQECIDRV 886
QY 631 NMYSSFYMMARR 641
Db 887 NMYASCMILEK 897

RESULT 4

US-10-118-495-3
; Sequence 3, Application US/10118495
; Publication No. US20030074688A1
; GENERAL INFORMATION:
; APPLICANT: Benning, Christoph
; APPLICANT: Riekhof, Wayne
; TITLE OF INVENTION: Compositions and Methods for the Production of Betaine Lipids
; CURRENT APPLICATION NUMBER: US/10/118,495
; CURRENT FILING DATE: 2002-04-08
; PRIOR APPLICATION NUMBER: 60/283,812
; PRIOR FILING DATE: 2001-04-13
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 416
; TYPE: PRT
; ORGANISM: Rhodobacter sphaeroides
US-10-118-495-3

Query Match 11.0%; Score 379.5; DB 4; Length 416;
Best Local Similarity 28.1%; Pred. No. 8.3e-29;
Matches 127; Conservative 68; Mismatches 182; Indels 75; Gaps 20;
QY 231 TQGSIPVPLRPYVWIGRLPSVGHALH-----EERVPMPFPPTF---LYTQSW 280
Db 2 TQFALTHLP---APP---VAR--QIGAAVHRTSLLSAEGLMER--MFSRLFHGLVYPOIW 51
QY 281 EDPEPMEVMEINPKDVTLTSGGCNVALNLVQAGQVSVDCNPAQSALLEKKVAIQ 340
Db 52 EDPAVDMALAIRPGDLVAIASGGCNVLSYLTQGFSGTILAVDLSPAHVALGRLKLAAR 111
QY 341 QL-EFEDVWOLFGEVHPRIEELYEKKLAPFLSQTSHNF-----SKRLWYFQHGLYQ 393
Db 112 TLPDHAFFDLFGKADLPGNALYDRHIAPALDGRSRRYWEARSPFGRRIQLFERGYRH 171
QY 394 GGMGKLCWVLOCLAVLGLGKTVKRLANAPTMEEQRLWDSNMLIHFKVNGPKPLVWLFV 453
Db 172 GALGRFIGAHTLA---RAAGTDLRGFLDCPDIEAQRSPFFVAHI-----GP-----LFE 217
QY 454 KFSVSLVLFNKAVLWFGGVPKGQYALIKADG-----IPIENIYARTMDGVAENSH-----503
Db 218 APVQALARRPAALFGLGIPPAQYALLAGDGDVLPV-----LRQRLHLLCDF 267
QY 504 VRKQNYFYNNCLTGKFLRDN---CPTYLREAAAFATLKSGVVDNLTVSTNFFMELKARTY 560
Db 268 PIRENYFAQATARRPRPREGALPPYLEPTAFETLREN-AGRVQIENSLTALAAEPE 326
QY 561 TKV---ILMDHVDWLDMPVANELAECLAKOVA PGGIVWIR---SASLSPPYAEILQKAGF 614
Db 327 ESHGFTLLDAQDWMTDALQTLWRQVTRTAAPGARVIFRTGGAADLLP--GRVPBEILG 384
QY 615 DVRCIRATQG---YMDRVNMYSSFYMMARRKA 644
Db 172 GALGRFIGAHTLA---RAAGTDLRGFLDCPDIEAQRSPFFVAHI-----GP-----LFE 217

QY 454 KFSVSLVLFNKAVLWFGGVPKGQYALIKADG-----IPIENIYARTMDGVAENSH-----503
Db 218 APVQALARRPAALFGLGIPPAQYALLAGDGDVLPV-----LRQRLHLLCDF 267
QY 504 VRKQNYFYNNCLTGKFLRDN---CPTYLREAAAFATLKSGVVDNLTVSTNFFMELKARTY 560
Db 268 PIRENYFAQATARRPRPREGALPPYLEPTAFETLREN-AGRVQIENSLTALAAEPE 326
QY 561 TKV---ILMDHVDWLDMPVANELAECLAKOVA PGGIVWIR---SASLSPPYAEILQKAGF 614
Db 327 ESHGFTLLDAQDWMTDALQTLWRQVTRTAAPGARVIFRTGGAADLLP--GRVPBEILG 384
QY 615 DVRCIRATQG---YMDRVNMYSSFYMMARRKA 644
Db 385 HWRADRAAGAGHAADRSAYIGGFHLYRRDA 416

RESULT 5

US-10-620-914-3
; Sequence 3, Application US/10620914
; Publication No. US20040093639A1
; GENERAL INFORMATION:
; APPLICANT: Benning, Christoph
; APPLICANT: Riekhof, Wayne
; TITLE OF INVENTION: Compositions and Methods for the Production of Betaine Lipids
; FILE REFERENCE: MSU-07769
; CURRENT APPLICATION NUMBER: US/10/620,914
; CURRENT FILING DATE: 2003-07-16
; PRIOR APPLICATION NUMBER: 10/118,495
; PRIOR FILING DATE: 2002-04-08
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
; LENGTH: 416
; TYPE: PRT
; ORGANISM: Rhodobacter sphaeroides
US-10-620-914-3

Query Match 11.0%; Score 379.5; DB 4; Length 416;
Best Local Similarity 28.1%; Pred. No. 8.3e-29;
Matches 127; Conservative 68; Mismatches 182; Indels 75; Gaps 20;
QY 231 TQGSIPVPLRPYVWIGRLPSVGHALH-----EERVPMPFPPTF---LYTQSW 280
Db 2 TQFALTHLP---APP---VAR--QIGAAVHRTSLLSAEGLMER--MFSRLFHGLVYPOIW 51
QY 281 EDPEPMEVMEINPKDVTLTSGGCNVALNLVQAGQVSVDCNPAQSALLEKKVAIQ 340
Db 52 EDPAVDMALAIRPGDLVAIASGGCNVLSYLTQGFSGTILAVDLSPAHVALGRLKLAAR 111
QY 341 QL-EFEDVWOLFGEVHPRIEELYEKKLAPFLSQTSHNF-----SKRLWYFQHGLYQ 393
Db 112 TLPDHAFFDLFGKADLPGNALYDRHIAPALDGRSRRYWEARSPFGRRIQLFERGYRH 171
QY 394 GGMGKLCWVLOCLAVLGLGKTVKRLANAPTMEEQRLWDSNMLIHFKVNGPKPLVWLFV 453
Db 172 GALGRFIGAHTLA---RAAGTDLRGFLDCPDIEAQRSPFFVAHI-----GP-----LFE 217
QY 454 KFSVSLVLFNKAVLWFGGVPKGQYALIKADG-----IPIENIYARTMDGVAENSH-----503
Db 218 APVQALARRPAALFGLGIPPAQYALLAGDGDVLPV-----LRQRLHLLCDF 267
QY 504 VRKQNYFYNNCLTGKFLRDN---CPTYLREAAAFATLKSGVVDNLTVSTNFFMELKARTY 560
Db 268 PIRENYFAQATARRPRPREGALPPYLEPTAFETLREN-AGRVQIENSLTALAAEPE 326
QY 561 TKV---ILMDHVDWLDMPVANELAECLAKOVA PGGIVWIR---SASLSPPYAEILQKAGF 614
Db 327 ESHGFTLLDAQDWMTDALQTLWRQVTRTAAPGARVIFRTGGAADLLP--GRVPBEILG 384
QY 615 DVRCIRATQG---YMDRVNMYSSFYMMARRKA 644

```
Db 385 HWRADRAAGQAGHAADRSALYGGFHLVRRDA 416

RESULT 6
US-10-118-495-29
; Sequence 29, Application US/10118495
; Publication No. US20030074688A1
; GENERAL INFORMATION:
; APPLICANT: Benning, Christoph
; APPLICANT: Riekhof, Wayne
; APPLICANT: Klug, Rouven
; TITLE OF INVENTION: Compositions and Methods for the Production of Betaine Lipids
; FILE REFERENCE: MSU-06897
; CURRENT APPLICATION NUMBER: US/10/118,495
; CURRENT FILING DATE: 2002-04-08
; PRIOR APPLICATION NUMBER: 60/283,812
; PRIOR FILING DATE: 2001-04-13
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 29
; LENGTH: 416
; TYPE: PRT
; ORGANISM: Agrobacterium tumefaciens
US-10-118-495-29

Query Match 9.4%; Score 326.5; DB 4; Length 416;
Best Local Similarity 27.8%; Pred. No. 1.9e-23;
Matches 111; Conservative 69; Mismatches 158; Indels 61; Gaps 16;

QY 275 LYTQSWEDPEPDMVMEINPKDVTLTGSGCNALNLLVQAGQVSVDCNPAQSALLEL 334
Db 44 VYPIWEDPEIDMEAMELGEGHRIVITIGSGCNMLAYLSRNPASIDVVDLNPHHIALNKL 103
QY 335 KKVAIQOL-EFEDVWOLFEGVHPRIEELYEKLAFLPSQTSINFSK-----RLWYF 386
Db 104 KLAARHLPAHQDVVRHFGAGTRSNVSGVDRFIAEHLDTATTKAYWSKRTLSGRRRISVF 163
QY 387 QHGLYOGGMKLCWLQCLAVVLGLGKTVKRLANAPTMEEORRLWDSNNLIHFVKNGPK 446
Db 164 DRNIYRTGLGRFAGHIMARLHGKLT--ENAKRTLDEQRFDSKVAPLF-----DK 217
QY 447 PLV-WLFVKFVSLVFNKAVLMFGGVPQKQY---ALIKADGIPNIENYIARTMDGVAENS 502
Db 218 PVVRWLTTRKSSL-----FGLGIPPRQYDELASLSDG-TVASVLKERLEKLACNF 267
QY 503 HVKQNYFYNYCLTGKFLRDN---CPTYLREAAFAFLKSGVDNLTVSTNFFMEELK--- 556
Db 268 PL-SDNYFAWQAFARYPEPEHEGALPAYLKPEYKIRNNTA-RVAVHHATYTELLSRKP 325
QY 557 ARTVTVKILMDHVDLMDPVAELAECLAKQVAPGGIVWRSAS-----LSPPYAEL 608
Db 326 ANGVDRIYLLDAQDWMTDVQLNELWSQISRTASGARVIFRTAAEKSVIEGRISP----- 380
QY 609 IQKAGFDVR-----CIRRATQ-GYMDRVNMYSSFYMARR 641
Db 381 -----DIRNQWVLEERSNELNMDRSALYGGFHIYQR 413

RESULT 7
US-10-620-914-29
; Sequence 29, Application US/10620914
; Publication No. US2004009639A1
; GENERAL INFORMATION:
; APPLICANT: Benning, Christoph
; APPLICANT: Riekhof, Wayne
; APPLICANT: Klug, Rouven
; TITLE OF INVENTION: Compositions and Methods for the Production of Betaine Lipids
; FILE REFERENCE: MSU-07269
; CURRENT APPLICATION NUMBER: US/10/620,914
; CURRENT FILING DATE: 2003-07-16
; PRIOR APPLICATION NUMBER: 10/118,495
; PRIOR FILING DATE: 2002-04-08
; NUMBER OF SEQ ID NOS: 52

Db 385 HWRADRAAGQAGHAADRSALYGGFHLVRRDA 416

RESULT 8
US-10-118-495-41
; Sequence 41, Application US/10118495
; Publication No. US20030074688A1
; GENERAL INFORMATION:
; APPLICANT: Benning, Christoph
; APPLICANT: Riekhof, Wayne
; APPLICANT: Klug, Rouven
; TITLE OF INVENTION: Compositions and Methods for the Production of Betaine Lipids
; FILE REFERENCE: MSU-06897
; CURRENT APPLICATION NUMBER: US/10/118,495
; CURRENT FILING DATE: 2002-04-08
; PRIOR APPLICATION NUMBER: 60/283,812
; PRIOR FILING DATE: 2001-04-13
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 41
; LENGTH: 415
; TYPE: PRT
; ORGANISM: Mesorhizobium loti
US-10-118-495-41

Query Match 9.3%; Score 321; DB 4; Length 415;
Best Local Similarity 26.2%; Pred. No. 6.7e-23;
Matches 113; Conservative 70; Mismatches 176; Indels 72; Gaps 18;

QY 255 VGHAEERY-----ERPPMFPPTF---LYTQSWEDPEPDMVMEINPKDVTLTGSG 304
Db 16 VGVAVYQNRALSAGISER--LFAFLFSGLVYPIWEDPDVDMQGLQGHRIVTIASG 73
QY 305 GCNALNLLVQAGQVSVDCNPAQSALLELKKVIAIQOLEFE-DWQOLFEGVHPRIEELY 363
Db 74 GCNILAYLTRSPARIDAVDLNAAHIALNRMKLEAVRRLPSQGLFRFFGAADTSHNSQAY 133
```

QY 364 EKKLAPFLSQTSHNFWKRLW-----YFOHGLYYQGGMGKLCWLQCLAVVLG-----411
DB 134 DRFIAPHLDPVSRHYWERNRGRRIIAVDRNFYQTLGLG-----LFIAMGHRTAK 185
QY 412 -LGKTVKRLANAPTMEQORLWDSNMLIHFKVNGPKPLVWLFVKFVSLVLFNKAVLWFGG 470
DB 186 FFGVNPAMHWEARNIGEORFFNEELAPVFDK---KLLKWATSRKASL-----FGL 233
QY 471 GVPKQY--ALIKADGIPIENTYAR-----TMDGVAENSHVRKQNYFYNYNCLTGKFLRDN- 523
DB 234 GIPPAQYDSLITSGDGTMAVLKARLEKACDPLEN-----NYFAWQAFARRYPNCGE 287
QY 524 --CPTYLREAAFPATLKSQVVDNLTVSTNPFMBEL---KARTYTKVILMDHVDWLDMPVAN 578
DB 288 AALPAYLEKQNTETIR-GNIDRVAIHANLIEFLAGKADGTVDRLFLLDAQDMTDDQLN 346
QY 579 ELAECLAKQVAPGGIWIWSA---SLSP--PYAELIQKAGFVRCIRRTATOGVMDRVNMY 633
DB 347 ALWSEISRSTASAGARVIFRTAAEPSLLPGRVSTSLDDQWDYQDEASREFSA--RDRSAIY 404
QY 634 SSFYMARRKGA 644
DB 405 GGFHLYVKRTA 415

RESULT 9

US-10-620-914-41
; Sequence 41, Application US/10620914
; Publication No. US20040093639A1
; GENERAL INFORMATION:
; APPLICANT: Benning, Christoph
; APPLICANT: Riekhof, Wayne
; APPLICANT: Klug, Rouven
; TITLE OF INVENTION: Compositions and Methods for the Production of Betaine Lipids
; FILE REFERENCE: MSU-07769
; CURRENT APPLICATION NUMBER: US/10/620,914
; CURRENT FILING DATE: 2003-07-16
; PRIOR APPLICATION NUMBER: 10/118,495
; PRIOR FILING DATE: 2002-04-08
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 41
; LENGTH: 415
; TYPE: PRT
; ORGANISM: Mesorhizobium loti
US-10-620-914-41

Query Match 9.3%; Score 321; DB 4; Length 415;
Best Local Similarity 26.2%; Pred. No. 6.7e-23;
Matches 113; Conservative 70; Mismatches 176; Indels 72; Gaps 18;

QY 255 VGHALHEERV-----ERPPMPPTF---LYTQSWEDPEPDMVMEINPKDVTLTSG 304
DB 16 VGVKAVYONRALS KAGISER--LFAFLPSGLVYPOIWEEDPDVMEANQLCGQHRIIVTIASG 73
QY 305 GCNALMLLVQAGQVSDVCNPAQSALLLEKVAIQOLEFE--DVWQLFGEGVHPRIEELY 363
DB 74 GCNILAYLTRSPARIDAVDLNAAHIALNMKLEAVRPLSQGDLFFFGAADTSHNSQAY 133
QY 364 EKKLAPFLSQTSHNFWKRLW-----YFOHGLYYQGGMGKLCWLQCLAVVLG-----411
DB 134 DRFIAPHLDPVSRHYWERNRGRRIIAVDRNFYQTLGLG-----LFIAMGHRTAK 185
QY 412 -LGKTVKRLANAPTMEQORLWDSNMLIHFKVNGPKPLVWLFVKFVSLVLFNKAVLWFGG 470
DB 186 FFGVNPAMHWEARNIGEORFFNEELAPVFDK---KLLKWATSRKASL-----FGL 233
QY 471 GVPKQY--ALIKADGIPIENTYAR-----TMDGVAENSHVRKQNYFYNYNCLTGKFLRDN- 523
DB 234 GIPPAQYDSLITSGDGTMAVLKARLEKACDPLEN-----NYFAWQAFARRYPNCGE 287
QY 524 --CPTYLREAAFPATLKSQVVDNLTVSTNPFMBEL---KARTYTKVILMDHVDWLDMPVAN 578

DB 288 AALPAYLEKQNTETIR-GNIDRVAIHANLIEFLAGKADGTVDRLFLLDAQDMTDDQLN 346
QY 579 ELAECLAKQVAPGGIWIWSA---SLSP--PYAELIQKAGFVRCIRRTATOGVMDRVNMY 633
DB 347 ALWSEISRSTASAGARVIFRTAAEPSLLPGRVSTSLDDQWDYQDEASREFSA--RDRSAIY 404
QY 634 SSFYMARRKGA 644
DB 405 GGFHLYVKRTA 415

RESULT 10

US-10-118-495-33
; Sequence 33, Application US/10118495
; Publication No. US20030074688A1
; GENERAL INFORMATION:
; APPLICANT: Benning, Christoph
; APPLICANT: Riekhof, Wayne
; APPLICANT: Klug, Rouven
; TITLE OF INVENTION: Compositions and Methods for the Production of Betaine Lipids
; FILE REFERENCE: MSU-06897
; CURRENT APPLICATION NUMBER: US/10/118,495
; CURRENT FILING DATE: 2002-04-08
; PRIOR APPLICATION NUMBER: 60/283,812
; PRIOR FILING DATE: 2001-04-13
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33
; LENGTH: 416
; TYPE: PRT
; ORGANISM: Sinorhizobium meliloti
US-10-118-495-33

Query Match 8.4%; Score 292.5; DB 4; Length 416;

Best Local Similarity 26.0%; Pred. No. 5.1e-20;
Matches 102; Conservative 71; Mismatches 174; Indels 45; Gaps 16;

QY 275 LYTQSWEDPEPDMVMEINPKDVTLTSGGCNALMLLVQAGQVSDVCNPAQSALLLE 334
DB 45 VYPOIWEEDPDVMEANQIRPGHRIIVTIGSGGCNMLTYLSABPARIDVDLNPHHIALNRL 104
QY 335 KKVATQOL--EFEDVWQLFG--EGVHPRIEELYEKKLAPFLSQTSHNFW--KRLWY 385
DB 105 KLSAFRHLPSHKDVRFLAVEGTRTN--GOAYDVFLAPKLDPATRAYMNGRDLTGRRRIGV 163
QY 386 FOHGLYYQGGMGKLCWLQCLAVLGLKTVKRLANAPTMEQORLWDSNMLIHFKVNGP 445
DB 164 FGRNVTRTGLGRFITSASHALARLHGINP--EDFVKARSMRQORQFPDDKLAFLP---E 217
QY 446 KPLV-WLFVKFVSLVLFNKAVLWFGGVFGKQVALIKADGIPIENYIARTMDGVAENSHV 504
DB 218 RPVIRWITSRKSSL-----FGLGIPPOQDELAS--LSREKSVAAVLRNRLKLTJC 266
QY 505 R---KQNYFYNYNCLTGKFLRDN---CPTYLREAAFPATLKSQVVDNLTVSTNPFMBEL--- 555
DB 267 HFPLRDNYPAWQAFARRYPDEGELPPYLOASRYEARDN--ABREVHVHASFTTELAGK 325
QY 556 KARTYTKVILMDHVDWLDMPVANELAECLAKVAPGGIWIWS--ASLSPP--YAEIQL 610
DB 326 PAASVDVRYLLDAQDMTDDQLNDLWTEITRTADAGAVVIFRTAAEASILPGLRSLTLLD 385
QY 611 KAGFVRCIRRTATOGVMDRVNMYSSFYMARRK 642
DB 386 QWYDAETSMRL--GAEDRSALYGGPHIYRKK 415

RESULT 11

US-10-620-914-33
; Sequence 33, Application US/10620914
; Publication No. US20040093639A1
; GENERAL INFORMATION:
; APPLICANT: Benning, Christoph

; APPLICANT: Riekhof, Wayne
 ; APPLICANT: Klug, Rouven
 ; TITLE OF INVENTION: Compositions and Methods for the Production of Betaine Lipids
 ; FILE REFERENCE: MSU-07769
 ; CURRENT APPLICATION NUMBER: US/10/620,914
 ; CURRENT FILING DATE: 2003-07-16
 ; PRIOR APPLICATION NUMBER: 10/118,495
 ; PRIOR FILING DATE: 2002-04-08
 ; NUMBER OF SEQ ID NOS: 52
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 33
 ; LENGTH: 416
 ; TYPE: PRT
 ; ORGANISM: Sinorhizobium meliloti
 US-10-620-914-33

Query Match 8.4%; Score 292.5; DB 4; Length 416;
 Best Local Similarity 26.0%; Pred. No. 5.1e-20;
 Matches 102; Conservative 71; Mismatches 174; Indels 45; Gaps 16;
 QY 275 LYTSQWEDPEPDMVEINPKDVTILTSGGCNALLVQAGQVSVDCNPAQSALLEL 334
 DB 45 VYPQIWEDPVDMEANQIRPCHRVITIGSGCNMLTYLSAEPARIDVDLNPPIIALNRL 104
 QY 335 KKVAIOQL-EFEDVWOLFQ-EGVHPRIEELYKKLAPFLSQTSHNFWSS-----KRLWY 385
 DB 105 KLSAFPHLPSHKDQVDFLAVEGTRTN-GQAYDVFPLAPKLDPATRAYWNGEDLTGRRIGV 163
 QY 386 FOHGLYYQGGMGKLCWVLOCLAVVLGLGKTVKRLANAPTMEEQRLWDSNMLHFKVNGP 445
 DB 164 FGRNVVTRTGLLGRFISASHALARLHGPN-EDFVKARSREORQPFDDKLAPLF-----E 217
 QY 446 KPLV-WLFVKFVSLVLFNKAFLWFGGVGCKQVYALIKADGIPENTYIARTMDGVAENSHV 504
 DB 218 RPVIRWITRKSL-----FGLGIPQPFDELAS--LSREKSVAAVLRNRLKJLC 266
 QY 505 R---KQNYFYNCLTGKFLRDN---CPTVLRFAAFATLKGVDNLTSTNPFMEEL--- 555
 DB 267 HPLRDNYFAWQAFARYPRPDGELPPYLOASRYEARDN-AERVEVHHASFTLLAGK 325
 QY 556 KARTYTKVILMDHVDLMDPANELACLAKQVAPGIVWRS---ASLSPP--YALILQ 610
 DB 326 PAASVDRYLLDAQDWNWTDQQLNDLWTEITRTADAGAVVIFRTAAEASILPGLRLSTLLD 385
 QY 611 KAGFDVRCITRRATQGYMDRVNMYSSFYMARRK 642
 DB 386 QWYDAETSMRL--GAEDRSAYGGFHYRKK 415

RESULT 12
 US-10-684-141-58
 ; Sequence 58, Application US/10694141
 ; Publication No. US20050003536A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Furusawa, Mitsuru
 ; TITLE OF INVENTION: METHOD AND SYSTEM FOR RAPIDLY CONFERRING
 ; TITLE OF INVENTION: A DESIRED TRAIT TO AN ORGANISM
 ; FILE REFERENCE: 690116.401
 ; CURRENT APPLICATION NUMBER: US/10/684,141
 ; CURRENT FILING DATE: 2003-10-10
 ; NUMBER OF SEQ ID NOS: 66
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 58
 ; LENGTH: 2284
 ; TYPE: PRT
 ; ORGANISM: Mus musculus
 US-10-684-141-58
 Query Match 3.6%; Score 123; DB 5; Length 2284;
 Best Local Similarity 20.8%; Pred. No. 0.094;
 Matches 141; Conservative 79; Mismatches 217; Indels 240; Gaps 35;
 QY 107 HSLCEVAKKAKAK-GWKNQVQVVEADACQAPFPGTATLITFSYS----- 150

DB 422 HNL-----KAAAKAGLYDPVELDPEDMCRMA-TEQPTLATYVSDAVATYLYMKYVHP 476
 QY 151 -----LTMIPFFH-NVI-----DOACSYLSODGLGVGA 177
 DB 477 FIFALCTIIPMBPDEVLRKSGTLCALLMWQAFAHANIIFPNKQEQEFNKLJTDGHLVDA 536
 QY 178 DFVYSG-----KYDLP--LRQMPWSRFRFFWRSIFDIDNIDIGPERRAYLEQ--KL 223
 DB 537 EYVVGHVHVALSGVPRSDIPCFRPMNPAAPFOLLQORVEKTMRHAIEEBEKVPEQATNF 596
 QY 224 ERWEOQ--NTQGSIPYVPLWRAPYYVWIGRLPSVGHALHEERVERP-----PMFPPT 273
 DB 597 QEVCEQIKTKLTSKDVP-----NRIECPLIYHLDVGAMYVNI 634
 QY 274 FLYTQSWEDPEPDMVEINPKDVTILTSGGCN-----ALNLLVQAGQVSVDCN 325
 DB 635 ILTNR-----LQPSAIVDEATCAACDFNKPAGSCQCKQKMAWQWRGEFM----- 676
 QY 326 PAQSALLELKVAIOQLFEFEDVWOLFEGVHPRIEEL-----YEK-KLAPF----- 370
 DB 677 PASRS--EYHRIQ-HOLESEKFPPLFPEGPARAFHELSEEQAKYKRLADYCRKAYKK 733
 QY 371 -----LSQTSNFWKRLWYFQHGLYYQGGMGKLCWVLOCLAVVLGLGKTVKR 418
 DB 734 IHVTKVEERLTTICQRENSFYVDTVRAFRDRRYEPKGLHKWKKLSAAVEVGDASEVKR 793
 QY 419 LANAPTMESQRLWDSNMLH-----FVKNQPKPLVWLFVFKVSLVLFNKAVLWPF 468
 DB 794 CKN-----MEILYDSLQLAHKLILNSFYGYVMRKAR--WYSMEMAGIVCFT----- 838
 QY 469 GGGVPGCKQVYALIKADGIP-----ENYIARTMDGVAENSHVAKQNYFY 512
 DB 839 GANITQARELEIQGRPLELTDGTCWVLPNSFPENFIKT-----TNAKKPKLITSYP 893
 QY 513 NCLTGKFLRDNCTY-----LREAAFATLKGVDNLTSTNPFMEELKARTYTKVILMDH 568
 DB 894 GAMLINWKEGFTNHQYQELTEPSSLTY-----VTHSENSIFFEVDG-PYLAMIL--- 942
 QY 569 VDWLMDPANELACLAKQVAPGIVWRSASLSPYAEILQKAGFDVRCIRATQGYMD 628
 DB 943 -----PASKEGKKLKRYA-----VFNEGDSL-----AEL-----KGFEVK-----RRGELQ 980
 QY 629 RVNMY--SSFYMARRKGA 644
 DB 981 LIKIFOSSVFEAPLKGS 997

RESULT 13
 US-10-810-486-58
 ; Sequence 58, Application US/10810486
 ; Publication No. US20050054597A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Furusawa, Mitsuru
 ; TITLE OF INVENTION: METHOD AND SYSTEM FOR RAPIDLY CONFERRING
 ; TITLE OF INVENTION: A DESIRED TRAIT TO AN ORGANISM
 ; FILE REFERENCE: 690116.401C1
 ; CURRENT APPLICATION NUMBER: US/10/810,486
 ; CURRENT FILING DATE: 2004-03-26
 ; PRIOR APPLICATION NUMBER: US 10/684,141
 ; PRIOR FILING DATE: 2003-10-10
 ; PRIOR APPLICATION NUMBER: JP 2003-092898
 ; PRIOR FILING DATE: 2003-03-28
 ; NUMBER OF SEQ ID NOS: 95
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 58
 ; LENGTH: 2284
 ; TYPE: PRT
 ; ORGANISM: Mus musculus
 US-10-810-486-58
 Query Match 3.6%; Score 123; DB 5; Length 2284;
 Best Local Similarity 20.8%; Pred. No. 0.094;

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 14, 2006, 02:01:27 ; Search time 9057 Seconds
(without alignments)
4066.971 Million cell updates/sec

Title: US-10-620-914-45
Perfect score: 3463
Sequence: 1 MSGRDRPASYTKNFKLSLE.....RVNMYSSFYMARKGAKGN 648

Scoring table:

BLOSUM62	Xgapop 10.0 , Xgapext 0.5
-Q-/abes/ABSSWEB	Ygapop 10.0 , Ygapext 0.5
-Q-/abes/ABSSWEB	Ygapop 6.0 , Ygapext 7.0
Delop 6.0 , Delext 7.0	

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

```
-MODEL=frame+ p2n.model -DEV=xlh
-Q-/abes/ABSSWEB spool/US10620914/runat_13032006_102040_6725/app_query.fasta_1
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdd -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abes05h
-USER=US10620914 @CGN 1.1 4939 @runat_13032006_102040_6725 -NCPU=6 -ICPU=3
-NO MWAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7
```

Database :

```
GenEmbl.*
1: gb.ba.*
2: gb.in.*
3: gb.env.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pr.*
9: gb.ro.*
10: gb.sts.*
11: gb.sy.*
12: gb.un.*
13: gb.vi.*
14: gb.htg.*
15: gb.pl.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	3441	99.4	2001	15	AY656806	Chlamydom
2	913.5	26.4	31065	15	AC152119	Pichia st
3	910	26.3	110000	15	CR382131_12	Continuation (13 o

	4	890.5	25.7	110000	15	CR382126_20	Continuation (21 o
c	5	864.5	25.0	110000	15	CR382138_20	Continuation (21 o
	6	804	23.2	75337	15	AX842626	Neurospor
	7	724	20.9	6962	15	AJ878596	AJ878596 Rhizopus
	8	633	18.3	110000	15	AE017345_05	Continuation (6 of
c	9	460	13.3	311050	1	AX294133	BY294133 Pirellula
	10	395	11.4	349260	1	AX572595	AX572595 Rhodosphe
	11	379.5	11.0	3045	1	AF329857	AF329857 Rhodobact
	12	352	10.2	933	6	AE551999	AE551999 Sequence
	13	336.5	9.7	10129	1	AE008128	AE008128 Agrobacte
	14	336.5	9.7	10134	1	AE009162	AE009162 Agrobacte
	15	329	9.5	110000	1	BA000012_13	Continuation (14 o
	16	292.5	8.4	294800	1	SME591789	ALS591789 Sinorhizo
c	17	277	8.0	86289	14	CR954215_4	Continuation (5 of
c	18	243.5	7.0	756	10	PM8H7B	AL685426 Penicilli
c	19	232.5	6.7	661	10	PM11C4B	AL684149 Penicilli
c	20	178	5.1	344249	1	AX842654	AX842654 Bdeellovib
c	21	159	4.6	110000	1	AP006618_16	Continuation (17 o
c	22	158	4.6	866	6	AR557101	AR557101 Sequence
c	23	138.5	4.0	8832	6	AR619589	AR619589 Sequence
c	24	137	4.0	725	10	PM3B11G	AL684886 Penicilli
c	25	132.5	3.8	306358	1	AE016759	AE016759 Escherich
c	26	131.5	3.8	300409	1	AE016755	AE016755 Escherich
c	27	128	3.7	4246	6	AX702426	AX702426 Sequence
c	28	126.5	3.7	590	6	AR627802	AR627802 Sequence
c	29	123.5	3.6	203961	8	AC011374	AC011374 Homo sapi
c	30	123	3.6	7119	6	CQ893379	CQ893379 Sequence
c	31	123	3.6	7119	9	AF123502	AF123502 Mus muscu
c	32	123	3.6	340900	1	SME591791	ALS591791 Sinorhizo
c	33	122	3.5	1604	5	CC515KDP	X15825 Quail mRNA
c	34	122	3.5	7102	9	BC048166	BC048166 Mus muscu
c	35	122	3.5	7102	9	BC063246	BC063246 Mus muscu
c	36	122	3.5	7954	1	AE010932	AE010932 Methanosa
c	37	121	3.5	639	6	AR348989	AR348989 Sequence
c	38	121	3.5	10029	1	AE013909	AE013909 Yersinia
c	39	121	3.5	210050	1	AJ414146	AJ414146 Yersinia
c	40	121	3.5	290510	1	AE017130	AE017130 Yersinia
c	41	119	3.4	5067	8	AB209902	AB209902 Homo sapi
c	42	119	3.4	6912	6	CQ893375	CQ893375 Sequence
c	43	119	3.4	6912	8	HUMDNPOLCS	L09561 Homo sapien
c	44	118.5	3.4	1155	6	BD242271	BD242271 Compounds
c	45	118.5	3.4	1155	6	AR278478	AR278478 Sequence

ALIGNMENTS

RESULT 1

AY656806

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

AY656806 2001 bp mRNA linear PLN 11-FEB-2005
Chlamydomonas reinhardtii betaine lipid synthase (BTAL) mRNA,
complete cds.

AY656806 GI:50261571

Chlamydomonas reinhardtii

Chlamydomonas reinhardtii

Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;

Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.

1 (bases 1 to 2001)

Riekhof, W.R., Sears, B.B. and Benning, C.

Annotation of Genes Involved in Glycerolipid Biosynthesis in

Chlamydomonas reinhardtii: Discovery of the Betaine Lipid Synthase

BTALCR

Eukaryotic Cell 4 (2), 242-252 (2005)

15701786

2 (bases 1 to 2001)

Riekhof, W.R. and Benning, C.

Direct Submission

Submitted (17-JUN-2004) BMB, MSU, Biochemistry 215, East Lansing,

MI 48824-1319, USA

Location/Qualifiers

1..2001

/organism="Chlamydomonas reinhardtii"

/mol_type="mRNA"		Db		481		ACGCTCATCACTTCTCTCTACTCGCTCAGATGATTCACACCGTTCACCAACGTCATCGAC		540	
/strain="CC125"		Qy		163		GlnAlaCysSerTyrLeuSerGlnAspGlyLeuValGlyValAlaAspPheTyrValSer		182	
/db_xref="taxon:3055"		Db		541		CAGGCTTGCTGCTACCTGTCCCAAGACGGCTGGTGGGCGTTCGCCACATTCACGTGAGC		600	
1. .2001		Qy		183		GlyLysTyrAspLeuProLeuArgGlnMetProTyrSerArgArgPhePheTyrArgSer		202	
/genes="BTA1"		Db		601		GGCAAGTAGCACTGCCCTCTGCCAGATGCCCTGTGTGGCGCGTTCCTTCTGGCGATCG		660	
/notes="BTA1"		Qy		203		IlePheAspIleAspAsnIleAspIleGlyProGluArgAlaTyrLeuGluGlnLys		222	
S-adenosylmethionine:diacylglycerol		Db		661		ATCTTCGACATCGACAACATTGACATCGGCCCGCGCGCGCTACCTCGAGCAGAG		720	
S-adenosylmethionine:diacylglycerolhomoserine		Qy		223		LeuGluArgValTyrGluGlnAsnThrGlnGlySerIleProTyrValProTyrLeuArg		242	
N-methyltransferase activities"		Db		721		CTGGAGCGGTGTGGGAGCAGAACACCCAGGGTTCGATCCCCCTACGTGGCGTGGCGGC		780	
/codon_start=1		Qy		243		AlaProTyrTyrValTyrIleGlyArgLeuProSerValGlyHisAlaLeuHisGluGlu		262	
/product="betaine lipid synthase"		Db		781		GCCCCCTACTAGTGTGGATTGGCCGCTGCCAGGTTGGCCACGCCCTGCACAGGAG		840	
/protein_id="AA172342.1"		Qy		263		ArgValGluArgProMetPheProTyrPheLeuTyrThrGlnSerTyrGluAsp		282	
/db_xref="GI:50261572"		Db		841		CGCGTGGAGCGCGCCCATGTTCCCGCCACCTCTCTGTACACGCGAGTCTGTGGAGGAC		900	
/translations="MGSGRDGRPASYYTKRNFSLKLSJSMKDDLTVLRLHMFSGKKG		Qy		283		ProGluProAspMetGluValMetGluIleAsnProLysAspThrValLeuThrLeuThr		302	
DHAALLESFYGPQAAAYDAFRSRFLWGRPMLAAVAARLAERSNLIWDLGGGTGEN		Db		901		CCCGAGCCGATATGGAGGTGATGGAGATCAACCCCAAGGACACGGTGTGACCCCTGACT		960	
VDMWADYIDLAKFPKSYVVDLCHSLCEVAKKAKAGKNVVOVEADACQFAPPEGTA		Qy		303		SerGlyGlyCysAsnAlaLeuAsnLeuLeuValGlnGlyValGlnValValSerVal		322	
TLTFYSLSLTMIPFNHNDACSYLSQDLGVADFYVSGKYVDLPLRMPWSRRFFW		Db		961		AGCGGCGGTGCTGATGCTTGAACCTGTGTGGTGCAGGGGCGGCCGAGGTGGTGGTG		1020	
LHEERVERPMPFTTYTQSWEDPEPDMVEINPKDTVLTLTSGCCNALNLVQGA		Qy		323		AspCysAsnProAlaGlnSerAlaLeuLeuGluLeuLysLysValAlaIleGlnGlnLeu		342	
GVVSDPCNPAQSLLEKLVAIQOIEFEDVQLFGEGVHPRIEELYKKLAPFLSOT		Db		1021		GACTGCAACCCCGCGCAGTCGGCGCTTCTGGAGTCAAGAAGGTGGCCATTCAGCAGCTG		1080	
SHNFWKRLWYFQGLYQYOGMGLCWLVCLAVLGLGKTVKRLANATMBEQRLW		Qy		343		GluPheGluAspValTyrGlnLeuPheGlyGlyValHisProArgIleGluGluLeu		362	
DSNMLTHFKVNGPKPLVLFVSVLFNKAVLWFGGVPKGQYALIKADGPIENY		Db		1081		GAGTGTGGAGACGTGTGGCAGCTGTTCGGCGAGGGCGTGACCCGCGCATTTGAGGAGCTG		1140	
IARTMDGVAENSHVRQNYFYNCLTGKPLRDNCPYLRBAAPATLKSQVNDLTVST		Qy		363		TyrGluLysLysLeuAlaProPheLeuSerGlnThrSerHisAsnPheTyrSerLysArg		382	
		Db		1141		TACGAGAGAGCTGGCGCCCTTCTGTGCAAAACACGACCACTTCTGTGTCACAGCGC		1200	
NFMEELKARTYTKYILMDHVDMDMPVANELABCLAKQVAPGGIVLWRSASLSPPYA		Qy		383		LeuTyrTyrPheGlnHisGlyLeuTyrTyrGlnGlyMetGlyLysLeuCysTyrVal		402	
ELIOKAGFDVRCIRRTATQGYMDRVNMYSPFMARRKGAKKDN"		Db		1201		CTCTGGTACTTCCAGCACGCGCTGTACTACAGGGCGGCATGGCAAGCTGTGTGGGTG		1260	
US-10-620-914-45 (1-648) x AY656806 (1-2001)		Qy		403		LeuGlnCysLeuAlaValValLeuGlyLeuGlyLysThrValLysArgLeuAlaAsnAla		422	
1 MetGlySerGlyArgAspGlyArgProAlaSerTyrThrLysLysAsnPheSerLeuGlu		Db		1261		CTGCAGTGCCTGGCGGTGTGGAGCTGGGCAAGACCGTCAAGCGCTCGCCACAGCG		1320	
1 ATGGGGTCGGGTCTGTGAGCGCCGCGCTCGAGCTACACCAAGAAGAACTTCTCCCTGGAG		Qy		423		ProThrMetGluGlnArgArgLeuTyrAspSerAsnMetLeuIleHisPheValLys		442	
21 LysLeuLysLeuSerSerMetLysAspAspLeuThrValLeuArgHisMetTyrPheGly		Db		1321		CCACACATGGAGAGCAGCGCGCTGTGTGGACAGCAACATGTCTCATCTTCGTGAAG		1380	
61 AAGCTCAAGCTCAGCAGATGAAGGATGACCTGACCGTTCGCGCCATATGTGTGGCG		Qy		443		AsnGlyProLysProLeuValTyrPheValLysPheValSerLeuValLeuPheAsn		462	
41 SerLysLysGlyAspAspHisAlaAlaArgLeuGluSerPheTyrGlyProGlnAlaAla		Db		1381		AACGGGCCCCAAGCGCTGTGTGGCTTCTCGTCAAGTTCGTGAGCTGTGTCTTCAAC		1440	
121 AGCAAGAAGGCGGATGATACGCTGCTGCGCTGGAGAGCTTCTACGGGCCCCAGCGCT		Qy		463		LysAlaValLeuTyrPheGlyGlyValProGlyLysGlnTyrAlaLeuIleLysAla		482	
61 AlaPhe-----		Db		1441		ARGCGCGTGTGTGTGGTTCGGCGCGCGCTGCCGGGCAAGCAGTACGGCGCTGATCAAGCG		1500	
181 GCCTATGATGTTTCGGTTCGGGTCTCTCTGGGGTTCGAGGCCCATGCTCGTCAGTT		Qy		483		AspGlyIleProIleGluAsnTyrIleAlaArgThrMetAspGlyValAlaGluAsnSer		502	
63 AlaAlaArgLeuAlaGluArgSerAsnLeuIleTyrValAspLeuGlyGlyThrGly		Db		1501		GACGGCATCCCATTTGAGAACTACATCGCGCGCACCATGACGCGGTGGCGGAACTCG		1560	
241 GCTGCCCGCTGGCGAGCGCTCGAACCTCATCTGGGTGTGACCTGGGTGTGGCACTGGG		Qy		503		HisValArgLysGlnAsnTyrPheTyrTyrAsnCysLeuThrGlyLysPheLeuArgAsp		522	
83 GluAsnValAspMetMetAlaAspTyrIleAspLeuAlaLysPheLysSerIleTyrVal		Db							
301 GAGNATGTCATATGATGCTGATTATCATCGACCTGGCGAGTTCAGTCCATCTAGTG		Qy							
103 ValAspLeuCysHisSerLeuGluValAlaLysLysLysAlaLysAlaLysGlyTyr		Db							
361 GTCGACCTGTGCCACTCGCTGTGGCGAGTGGCCCAAGAAGAAGGCGCAAGGGCTGG		Qy							
123 LysAsnValGlnValValGluAlaAspAlaCysGlnPheAlaProProGluGlyThrAla		Db							
421 AAGAATGTCCAGGTCTGTGGAGGCGCGCTGTGCAATTTGCGCCCTCGAGGGGACCGCG		Qy							
143 ThrLeuIleThrPheSerTyrSerLeuThrMetIleProPheHisAsnValIleAsp		Db							

ORIGIN

Alignment Scores:

Pred. No.: 5,1e-296 Length: 2001
 Score: 3441.00 Matches: 647
 Percent Similarity: 97.3% Conservative: 1
 Best Local Similarity: 97.1% Mismatches: 0
 Query Match: 99.4% Indels: 18
 DB: 15 Gaps: 1

```

Db      1561  CACGTGGCAGCAGCACTACTTCTACTACAACTGCTCACCGCAAGTTCCTGCGGAC 1620
Qy      523  AsnCysProThrTyrLeuArgGluAlaAAlaPheAlaThrLeuLysSerGlyValValAsp 542
Db      1621  AACTGCCCACTTACCTTGGCGGAGCGGCTTCCGCCACCTCAAGAGTGGCGTGGAC 1680
Qy      543  AsnLeuThrValSerThrAsnPheMetGluGluLeuLysAlaArgThrTyrThrLys 562
Db      1681  AACCTGACCGCTCCCACTTCTTCATGGAGGCTCAAGCGCGCACTTACACCAAG 1740
Qy      563  ValLeuLeuMetAspHisValAspTyrLeuAspMetProValAlaAsnGluLeuAlaGlu 582
Db      1741  GTGATCTGTATGACCACTGACTGGCTGGTATATGCCCGTGGCCACGAGCTGGCGAG 1800
Qy      583  CysLeuAlaLysGlnValAlaProGlyGlyLeuValLeuTyrArgSerAlaSerLeuSer 602
Db      1801  TGCCTGGCCAAAGCAGGTGGCGCGGCGGATCGTATCTGGCGCTCCGCTCCCTCAGC 1860
Qy      603  ProProTyrAlaGluLeuLeuGlnLysAlaGlyPheAspValArgCysIleArgArgAla 622
Db      1861  CCGCCCTACCGCAGCTGATCCAGAGGGCGGCTTCGACGTGGCTGCATCCGCGCGCC 1920
Qy      623  ThrGlnGlyTyrMetAspArgValAsnMetTyrSerSerPheTyrMetAlaArgArgLys 642
Db      1921  ACTCAGGCTACATGACCGCTCAACATGTACAGCTCTTCTACATGGCCCGCGAAG 1980
Qy      643  GlyAlaLysLysAspAsn 648
Db      1981  GCGGCCAAGAAGGACAAC 1998

RESULT 2
AC152119 31065 bp DNA linear PLN 30-NOV-2004
LOCUS Pichia stipitis clone JGIAHYZ-5A24, complete sequence.
DEFINITION AC152119
ACCESSION AC152119.2 GI:56121961
VERSION HTG.
KEYWORDS Pichia stipitis
SOURCE Pichia stipitis
ORGANISM Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Pichia.
REFERENCE 1 (bases 1 to 31065)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 31065)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (27-Oct-2004) Production Genomics Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA
94598-1698, USA
REFERENCE 3 (bases 1 to 31065)
AUTHORS Stanford Human Genome Center.
CONSTRM DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (30-NOV-2004) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On Nov 30, 2004 this sequence version replaced gi:54654153.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 100% of Sequence;
Estimated Total Number of Errors is 0.
FEATURES             Location/Qualifiers
     source            1..31065
                        /organism="Pichia stipitis"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:4924"
                        /clone="JGIAHYZ-5A24"
ORIGIN
Alignment Scores:

```

```

Pred. No.: 1,25e-69 Length: 31065
Score: 913.50 Matches: 223
Percent Similarity: 46.5% Conservative: 113
Best Local Similarity: 30.8% Mismatches: 244
Query Match: 26.4% Indels: 143
DB: 15 Gaps: 21

US-10-620-914-45 (1-648) x AC152119 (1-31065)
Qy      40  GlySerLysLysGlyAspAspHisAlaAlaArgLeuGluSerPheTyrGlyProGlnAla 59
Db      20013  GCAAGAAGAGGATCAGCAACATTCG-----CITGAGCAATTCCTACAAGAACAGGCT 20066
Qy      60  AlaAlaPhe----- 62
Db      20067  CACATCTACGACACACCACCGTGAGTCTTGTGTAAGGAGAGCAGGAATGTCTTCGTCTT 20126
Qy      63  ---AlaAlaArgLeuAlaGluArgSerAsnLeuLeuTyrValAspLeuGlyGlyGlyThr 81
Db      20127  GCTGCTCCATTTGCGAAGAGAAAGAGTGTGCTGATCGATGTGTTGGTGGAGTACT 20186
Qy      82  GlyGluAsnValAspMetMetAlaAspTyrIleAspLeuAlaLys---PheLysSerIle 100
Db      20187  GGTTCCAACATCGAGTACATGGACCATGCTCTCCATTCGAAACATTTAAGGCGGTC 20246
Qy      101  TyrValValAspLeuCysHisSerLeuGluValAlaLysLysLysLysAlaLys 120
Db      20247  TATTTGGTTGATTTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 20306
Qy      121  GlyTyrLysAsnValGlnValValGluAlaAspAlaCysGlnPheAlaProGluGly 140
Db      20307  AACTGAAAGAAATGTCAGGTCATTTGCTGATGCTGTGACTTTGAAATCGACTACAAG 20366
Qy      141  ThrAlaThrLeuIleThrPheSerTyrSerLeuThrMetIleProPheHisAsnVal 160
Db      20367  ACTGCTGACTTGATCAGTTTCTCTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 20426
Qy      161  IleAspGlnAlaCysSerTyrLeuSerGlnAspGlyLeuValGlyValAlaAspPheTyr 180
Db      20427  ATCGACATGCACTCACCAGTGTGGATAGCAGGGTATTATTGCTCTCTGTAGACTTTGGT 20486
Qy      181  ValSerGlyLysTyrAspLeuProLeu----- 189
Db      20487  ATCCAG---AGTTTGGACACCTCTATGGCGAGAAATCAACACTGTAGGAGGTTGGTCAAC 20543
Qy      190  ArgGlnMetProTyrSerArgPhePheTyrArgSerIlePheAspIleAspAsnIle 209
Db      20544  AGAATCATCCTCTGGTGTCTAGAACTTCTGGAGAAATTTGGTTGAAGCCGCAAGTC 20603
Qy      210  AspIleGlyProGluArgAlaTyrLeuGluGlnLysLeuGluArgValTyrGluGln 229
Db      20604  TTTTGGACTCATCGAAGAAATTAATTTGGAATACAAAGTTTGAACCACTCAAGTCTCTC 20663
Qy      230  AsnThrGln-----GlySerIleProTyrValProTyrLeuArgAlaProTyr 245
Db      20664  AACACCTATAACAAGAAAGTTGGTAAATC-----CCTTAC 20699
Qy      246  TyrValTyrIleGly-----ArgLeuProSerValGlyHisAlaLeuHisGlu 261
Db      20700  TACATCTGATTTGGCTGTGACCAAGTCCAAAGTCCGAAATATCTTACCGCTCTCACTGC 20759
Qy      262  GluArgValGluArgProMetPheProThr----- 273
Db      20760  TTGGCTACTGAATCTCTTACTTGGCACCTTCTACCACTCCCATTTACACTGCTCTCAA 20819
Qy      273  ----- 273
Db      20820  GACGTTCTTATCAGTAAGGGCCATGAAGCTGCGCTTGACGAACTTGCAAAAGAACTTACCT 20879
Qy      273  ----- 273
Db      20880  TTTCCTTCTTCTACTACCAGCGCGAATCTGGAGAGTCTACTATGAAGAGTTGAACCAG 20939

```

```
QY 274 -----PheLeuTyrThrGlnSerTrpGluAspProGluPro 285
Db 20940 CATTACGAACAGTTCAAGAACCAATACATCTACGCCTTCATCGGAGATCCACGTGAA 20999
QY 286 AspMetGluValMetGluIleAsnProLysAspThrValLeuThrLeuThrSerGlyGly 305
Db 21000 GATCACAATATCTCAAGTTCAACAGCAATGACACAGTTTTAGCCATTACTTCTGCGGGT 21059
QY 306 CysAsnAlaLeuAsn-----LeuLeuValGlnGlyAlaGlyGlnValValSerValAsp 323
Db 21060 GACAACACTTTATCTTATGTACTCTACCGGACCCCAAGAGAAATCCATGCTGTAGAC 21119
QY 324 CysAsnProAlaGlnSerAlaLeuLeuGluLeuLysLysValAlaIleGlnGlnLeuGlu 343
Db 21120 TTAACCCATCTCAAAACCACTTTGTGGAGTTGAAGTTGGCAGCTTTCAGATCTCTTGACA 21179
QY 344 PheGluAspValTrpGlnLeuPheGlyGlyValHisProArgIleGluGluLeuTyr 363
Db 21180 CACCAACAGATCTGGCAGATGTTGGTGAAGGTAAGATCGAAACTTCACCGACTTGTG 21239
QY 364 GluLysLysLeuAlaProPheLeuSerGlnThrSerHisAsnPheTrpSerLysArg--- 382
Db 21240 ATCGACAAGTTGTCCCAACATGCTCTTCCAAATGCCCTTCAGTACTGGATGCGACAGGA 21299
QY 383 ---LeuTrpTyrPheGlnHisGlyLeuTyrTyrGlnGlyGlyMetGlyLysLeuCysTrp 401
Db 21300 CACAAGTGTCTCCCAAGACGACTTAC-----CACACTGGTCTAGTAGTGG 21350
QY 402 ValLeuGlnCysLeuAlaVal-----ValLeuGlyLeuGlyLysThrValLysArg 418
Db 21351 GCTCTCAGATCGCCAAAGTACACTTTCAAGGCCTTTGGTGTCTACTAAAGATATCAAGCAG 21410
QY 419 LeuAlaAsnAlaProThrMetGluGluGlnArgLeuTrpAspSerAsnMetLeuIle 438
Db 21411 TTGTGCGAAGCAGCTACTATGGACGACAGATGAAGATCTCGGGAAGAGAGA----- 21461
QY 439 HisPheValLysAsnGlyProLysProLeuValTrpLeuPheValLysPheVal---Ser 457
Db 21462 -----GTGAAGAAGTCT-----TTGTTCACTCAGTAGTGGCTACT 21497
QY 458 LeuValLeuPheAsnLysAlaValLeuTrpPheGlyGlyValProGlyLysGlnTyr 477
Db 21498 TTATTGGTGGAAACCCACTCTTTTGTGGAAGCTCTTGAGAGTCCAGCTAATCAGGCC 21557
QY 478 AlaLeuIleLysAlaAspGlyLeuProIleGluAsnTyrIleAlaArgThrMetAspGly 497
Db 21558 TCCATGATG-----GGAGAATCCGCTCTCAAGTATGTTGTCGACAAATAGACCT 21608
QY 498 ValAlaGluAsnSerHisValArgLysGlnAsnTyrPheTyrTyrAsnCysLeuThrGly 517
Db 21609 CTCTTGAGAGATCTTGATTCAAGAGACAAATATTCTTACTACTTGTGCTTGATGAAC 21668
QY 518 LysPheLeuArgAspAsnCysProThrTyrLeuArgGluAlaAlaPheAlaThrLeu--- 536
Db 21669 AAGTACGCTACAAACAACCTGCTCCGACTCTGCTCAAGGCTACAAGAGATTGACG 21728
QY 537 -----LysSerGlyValValAspAsnLeuThrValSerThrAsnPhe 550
Db 21729 GCCCAGACTACTCCCAAGAGAAACCCATGGACAACATCAGAAATTCATCTGACTTT 21788
QY 551 PheMetGlu-----GluLeuLysAlaArgThrTyrThrLysValIleLeuMetAsp 567
Db 21789 TTGAACGACGTTGTCAGATGTCAGAGAGTCTGTTGACTATCGCAATATTATTGAGAC 21848
QY 568 HisValAspTrpLeu-----AspMetProValAlaAsnGluLeuAlaGluCysLeuAla 585
Db 21849 CACATGATTGTTGATGATGACTCCCAAGGATGTCACAGATGATGATTCGTTCTTAAG 21908
QY 586 LysGlnValAlaProGlyGlyValIleTrpArgSerAlaSerLeuSerProProTyr 605
Db 21909 GAAAGTCTTTCGACGGTGGTAGATCTGTGTGATGATCTGCTCTCAACCCCATGTAC 21968
QY 606 AlaGluLeuIleGlnLysAlaGlyPheAspValArgCysIleArg---ArgAlaThrGln 624
```

```
Db 21969 ATCAGCACCTTTGAGTCTATGGCTTTAGTTGTAAGCTGTCTGTCAGAATCTCAGGT 22028
QY 625 GlyTyrMetAspArgValAsnMetTyrSerSerPheTyrMetAlaArgArgLysGlyAla 644
Db 22029 GAATCTATTGACAGGGGTCAACATGTACGTCTTACTTGGGTATGCACCAAGTTCCAACC 22088
QY 645 LysLysAsp 647
Db 22089 AAGGAAGAC 22097
```

RESULT 3

CR382131_12

WPCOMMENT

Sequence split into 43 fragments LOCUS CR382131 Accession CR382131.

Fragment Name Begin End

```
CR382131_00 1 110000
CR382131_01 100001 210000
CR382131_02 200001 310000
CR382131_03 300001 410000
CR382131_04 400001 510000
CR382131_05 500001 610000
CR382131_06 600001 710000
CR382131_07 700001 810000
CR382131_08 800001 910000
CR382131_09 900001 1010000
CR382131_10 1000001 1110000
CR382131_11 1100001 1210000
CR382131_12 1200001 1310000
CR382131_13 1300001 1410000
CR382131_14 1400001 1510000
CR382131_15 1500001 1610000
CR382131_16 1600001 1710000
CR382131_17 1700001 1810000
CR382131_18 1800001 1910000
CR382131_19 1900001 2010000
CR382131_20 2000001 2110000
CR382131_21 2100001 2210000
CR382131_22 2200001 2310000
CR382131_23 2300001 2410000
CR382131_24 2400001 2510000
CR382131_25 2500001 2610000
CR382131_26 2600001 2710000
CR382131_27 2700001 2810000
CR382131_28 2800001 2910000
CR382131_29 2900001 3010000
CR382131_30 3000001 3110000
CR382131_31 3100001 3210000
CR382131_32 3200001 3310000
CR382131_33 3300001 3410000
CR382131_34 3400001 3510000
CR382131_35 3500001 3610000
CR382131_36 3600001 3710000
CR382131_37 3700001 3810000
CR382131_38 3800001 3910000
CR382131_39 3900001 4010000
CR382131_40 4000001 4110000
CR382131_41 4100001 4210000
CR382131_42 4200001 4224103
```

Continuation (13 of 43) of CR382131 from base 1200001 (CR382131 Yarrowia lipolytica chrom

Alignment Scores:

Pred. No.:	1-29e-68	Length:	110000
Score:	910.00	Matches:	226
Percent Similarity:	48.5%	Conservatives:	116
Best Local Similarity:	32.1%	Mismatches:	231
Query Match:	26.3%	Indels:	132
DB:	15	Gaps:	22

US-10-620-914-45 (1-648) x CR382131_12 (1-110000)

QY 51 LeuGluSerPheTyrGlyProGlnAla-----

Db 59694 CTGGAGTCTTCTTACAGGCGCCAGGCCAACATTATTCAGACCGAACCCGAGCCACTCTGTCTC 59753
Qy 60 -----AlaAlaPheAlaAlaArgLeuAlaGluArgSerAsnLeu 72
Db 59754 AAGGCGCAGAGCAGGCTCTGGCCACCACCGCCGCGAGCTTCGAAGAAGCATGATCTC 59813
Qy 73 IleTrrValAspLeuGlyGlyThrGlyGluAsnValAsp---MetMetAlaAspTyr 91
Db 59814 GTGTGGTTCACATTGAGGAGGAGAACCGGCTGGAAACATTGAGCACATGACCAACACTAC 59873
Qy 92 IleAspLeuAlaLysPheLysSerIleTyrValValAspLeuCyHisSerLeuCyGlu 111
Db 59874 CTTCCTCCCTCTCTTCAGGCGCATCTACTCTGTTGATCTGAGTCTCTCTCTGTGGAG 59933
Qy 112 ValAlaLysLysAlaLysAlaLysGlyTrrLysAsnValGlnValValGluAlaAsp 131
Db 59934 GTGGCCGAAGCGATTTCGCCGAGAGAACTGGAGAACTGTCACGTTCTGTGTCATCGAC 59993
Qy 132 AlaCysGlnPheAlaProProGluGlyThrAlaThr---LeuIleThrPheSerTyrSer 150
Db 59994 GCTGCCGATTCTACTCCCGCGTGGATCTCGGTTGATCTCTTCCACATGCTCTACTCT 60053
Qy 151 LeuThrMetIleProProPheHisAsnValIleAspGlnAlaCysSerTyrLeuSerGln 170
Db 60054 CTGTCTATGATCCCACTACTACTATGCGTTATCGATCGACTCTCCACCTGCTTGGCAAG 60113
Qy 171 AspGlyLeuValGlyValAlaAspPheTyrValSerGlyLysTyrAspLeuPro----- 188
Db 60114 GACGGTCTGTCTACTGTCATCGACTTTTACGTCCTCAATCCATCCACCTTGACCGCCAG 60173
Qy 189 -----LeuArgGlnMetProTrrSerArgPhePheTrr 200
Db 60174 TCTACTACATGGCGCGAGCTGTCTCCGCGATGTCAACTGGTTCTCCCGAACTTTCTGG 60233
Qy 201 ArgSerIlePheAspIleAspAsnIleAspIleGlyProGluArgAlaTyrLeuGlu 220
Db 60234 CGACTGTGGTTCGAGTTGACCGGGTCTACTCGACTCTGCCGAGAGACTATCTCGAG 60293
Qy 221 GlnLys-----LeuGluArgValTrrGluGlnAsnThrGln---GlySerIlePro 236
Db 60294 TACCGATTCCGAACCATCAGTCAGTCAGTCGCGGAACACCGACTCGGAGGAACTCCCC 60353
Qy 237 TyrValProTrrLeu----- 241
Db 60354 TACTACTACTGTGATTGGATGTGACAAAGACCGATCCACCAACATTCTCCAGCGAGCCAAC 60413
Qy 242 -----ArgAlaProTrrTyr-----ValTrrIleGlyArgLeu 252
Db 60414 GCTCTGTGTACCGAGTCTCCCTACCTGTCTCCCGCAGAACAGAGCTTGTGTGGAGACGAG 60473
Qy 253 ProSerValGlyHisAlaLeu----- 259
Db 60474 GACGCGCTAACACACTCTGGGTATCCGATCCAGGGCTACGATGCCGCTCTCGTCAAC 60533
Qy 259 ----- 259
Db 60534 ATGCAGCGAACTTCCCGCTCCCTCATCTTCTTACAGACCGAAATCTGGCGAACTTCTC 60593
Qy 260 HisGluGluArgValGluArgProProMetPheProThrPheLeuTyrThrGlnSer 279
Db 60594 TACAACAGGAGTGTGCCAGTAGTACCACAGTTTGGCCAACAGTACATCTACGCTTCCACC 60653
Qy 280 TrrGluAspProGluProAspMetGluValMetGluIleAsnProLysAspThrValLeu 299
Db 60654 TGGGAGGACCCCGAGGAGACAGAACTTCTACAGTTTAAGCCCGACGATACCGTTCTT 60713
Qy 300 ThrLeuThrSerGlyGlyCysAsnAlaLeuAsnLeuValGlnGlyAla-----Gly 317
Db 60714 GCCATCAGCTCTGCGGTGACAACTTCTGTCTTACGCTCCCATGAGCGCTCCCGCCAG 60773
Qy 318 GlnValValSerValAspCysAsnProAlaGlnSerAlaLeuLeuLysLysVal 337
Db 60774 CGAATCCATTGTGCTCAACCTTCCAGAACCATCTGCTCGAGCTCAAGCTGGCC 60833

RESULT 4

CR382126_20

WPCOMMENT

Sequence split into 26 fragments LOCUS CR382126 Accession CR382126

Qy 338 AlaIleGlnGlnLeuGluPheGluAspValTrrGlnLeuPheGlyGluGlyValHisPro 357
Db 60834 TGTCTCGAGTGTCTTCCCTTTGAGGACATGTGTGAAGCTGTTTGTGAAGGCAAGCACCCC 60893
Qy 358 ArgIleGluGlnLeuTyrGluLysLysLeuAlaProPheLeuSerGlnThrSerHisAsn 377
Db 60894 AAGTTCGAAGAGTGTGTGACCAAGCTGGCCCCCACCCTGTCTTCCCAAGCTTCCAG 60953
Qy 378 PheTrrSerLysArgLeuTrrPheGlnHisGlyLeuTyrTyrGlnGlyGlyMet--- 396
Db 60954 TACTGSCACAGCGA-----GCCACGCTCTTTCACCGAAGGCTGTCTTT 61001
Qy 397 -----GlyLysLeuCysTrrValLeuGlnCys-----LeuAlaValValLeuGly 411
Db 61002 GATACCGGTCTTCCCGATGGGCCATTTCGATTTGCCATTGGGTCTTGTGCAATTTCTGTT 61061
Qy 412 LeuGlyLysThrValLysArgLeuAlaAsnAlaProThrMetGluGlnGlnArgArgLeu 431
Db 61062 GTCTCCCCCACAATTAGCGCTCTGTGCGAGGCCAAGACTCTCGACGAACAGTGGAGCGTC 61121
Qy 432 TrrAspSerAsnMetLeuIleHisPheValLysAsnGlyPro-----LysProLeu 448
Db 61122 TGGGAGAGTCCCTT-----CGACCTGTCTGTCCAAACCCCATTT 61169
Qy 449 ValTrrPhePheValLysPheValSerLeuValLeuPheAsnLysAlaValLeuTrrPhe 468
Db 61161 GTC-----GCCAAGGTGTGTTGGCAACCCCATTTTCTTGTGTGAAG 61202
Qy 469 GlyGlyGlyValProGlyLysGlnTyrAlaLeuIleLysAlaAspGlyIleProIleGlu 488
Db 61203 GCTCTCGGTGTTTCCCTGTGACGACCTCTATGATCGAGGCGGTATGCTC----- 61253
Qy 489 AsnTrrIleAlaArgThrMetAspGlyValAlaGluAsnSerHisValArgLysGlnAsn 508
Db 61254 AAGTTTGTCTACGACACTTCGAGCCCATCATCAGCGATCGCTCATTTCCGACGACAC 61313
Qy 509 TyrPheTyrTyrAsnCysLeuThrGlyLysPheLeuArgAspAsnCysProThrTyrLeu 528
Db 61314 TACTTCTACTCTGTGTCTCAAGGCTGTCTAGCGTCCCAACAACTGTCCGACTTCTCTG 61373
Qy 529 ArgGluAlaAlaPheAlaThrLeu-----LysSerGlyValValAspAsnLeuThrVal 546
Db 61374 ACCAAGCAGCGCCACCCCACTGGCCAAAGCGAAGGAGCTCTTGACGGAATCCGAATC 61433
Qy 547 SerThr-----AsnPhePheMetGluLeuLysAlaArgThrTyrThrLysVal 563
Db 61434 CACACCGAGATCAACGAGGTGGTCAAGCGACTCAACCCGATCAGTCAACACGCC 61493
Qy 564 IleLeuMetAspHisValAspTrrLeu-----AspMetProValAlaAsnGluLeuAla 581
Db 61494 ATCGTTATGGACCATGAGTGTCTCCCAAGGACGGTAACGCCGCTCGAGAGGAGATC 61553
Qy 582 GluCysLeuAlaLysGlnValAlaProGlyGlyIleValIleTrrPargSerAlaSerLeu 601
Db 61554 AAGTCTCTCCACCATGCTCTTACCGAAGGTGAAACGTTATGTTCGATCAGCTCCAC 61613
Qy 602 SerProTrrTrrAlaGluLeuIleGlnLysAlaGlyPheAspValArg-----CysIle 619
Db 61614 AAGCCCTGTGTACTCTAGCGTCTATGAGGAGGAGGCTTACCACTAGACCCGCTGCCATC 61673
Qy 620 ArgArgAlaThrGlnGlyTyrMetAspArgValAsnMetTyrSerSerPheTyrMetAla 639
Db 61674 CGAGAATCTGGCACTTCC---ATTGACCGAGTCAACATGTACGCTTCCACTTGGTCTGT 61730
Qy 640 ArgArgLysGlyAla 644
Db 61731 ACCAAGAGGAGGCC 61745

Fragment Name	Begin	End	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
CR382126_00	1	110000	7,01e-67	890,50	46.6%	33.2%	25.7%	15
CR382126_01	100001	210000						
CR382126_02	200001	310000						
CR382126_03	300001	410000						
CR382126_04	400001	510000						
CR382126_05	500001	610000						
CR382126_06	600001	710000						
CR382126_07	700001	810000						
CR382126_08	800001	910000						
CR382126_09	900001	1010000						
CR382126_10	1000001	1110000						
CR382126_11	1100001	1210000						
CR382126_12	1200001	1310000						
CR382126_13	1300001	1410000						
CR382126_14	1400001	1510000						
CR382126_15	1500001	1610000						
CR382126_16	1600001	1710000						
CR382126_17	1700001	1810000						
CR382126_18	1800001	1910000						
CR382126_19	1900001	2010000						
CR382126_20	2000001	2110000						
CR382126_21	2100001	2210000						
CR382126_22	2200001	2310000						
CR382126_23	2300001	2410000						
CR382126_24	2400001	2510000						
CR382126_25	2500001	2602197						
Continuation (21 of 26) from base 2000001 (CR382126_Kluyveromyces lactis str								
Alignment Scores:								
Pred. No.:	7,01e-67	Length:	110000					
Score:	890,50	Matches:	233					
Percent Similarity:	46.6%	Conservative:	94					
Best Local Similarity:	33.2%	Mismatches:	241					
Query Match:	25.7%	Indels:	133					
DB:	15	Gaps:	23					
US-10-620-914-45 (1-648) x CR382126_20 (1-110000)								
Qy	51	LeuGluSerPheTyrGlyProGlnAlaAlaPhe-----	62					
Db	75238	TTAGAACAAATTTATAAAATCCCAAGCCAACTCTACGATCGTACAAAGGGGTGTTTACTC	75297					
Qy	63	-----AlaAlaArgLeuAlaGluArgSerAsnLeu	72					
Db	75298	CAAGTCGTGAACAAAGCCTTAATATCTCTTCTCACCTCTCTGAAGAAAGAGTAAT	75357					
Qy	73	IleTrpValAspLeuGlyGlyThrGlyGluAsnValAspMetMetAlaAspTyrIle	92					
Db	75358	GTCTGGATTGATGCGGGGTGGGACTGGTTTCAATATCTCCCAATGGCGCTTTTAAACA	75417					
Qy	93	AspLeu---AlaLysPheLysSerIleTyrValValAspLeuCysHisSerLeuCysGlu	111					
Db	75418	AACTTGGATACACAGTTTGTGACAAAATTTACTTGTGATGATCTCTCCATCGCTATGTGAG	75477					
Qy	112	ValAlaLysLysLysAlaLysAlaLysGlyTrpLysAsnValGlnValValGluAlaAsp	131					
Db	75478	GTTCAGAAAGACCTTGAAGACATGGATGGAAATATGGGAAGTGAICTGTGGAGAT	75537					
Qy	132	AlaCysGlnPheAlaProProGluGlyThrAlaThrLeuIleThrPheSerTyrSerLeu	151					
Db	75538	GCTTGCATTTTGAATATACACAGAGAAATCAGTCAGTTGATCACAATTTCTTATTCATTG	75597					
Qy	152	ThrMetIleProProPheHisAsnValIleAspGlnAlaCysSerTyrLeu---SerGln	170					
Db	75598	AGCATGATACCAAGTTTCTTCGCGCCATTTGATCAGCAGTATCTTTTATTGGATGCTAAG	75657					
Qy	171	AspGlyLeuValGlyValAlaAspPheTyrValSerGlyLysTyr-----	185					
Db	75658	AATGGTATCATTTCTGTGTGATTTTGGTGTCTACTAATGAAATCTATGCTAGTCGGAAGA	75717					
Qy	186	-----AspLeuProLeuArgGlnMetProTrpSerArgPhePheTrp	200					

Db	75718	ACCAACACATTTGGTGGTCTAGTTAAACAGACATATACCTTTGGCTATTTCGTACATTCTGG	75777					
Qy	201	ArgSerIlePheAspIleAspAsnIleAspIleGlyProGluArgArgAlaTyrLeuGlu	220					
Db	75778	AGATTGTGGTTTGAATTCGACAAAGTGTGTTTGGATCCAGCTAGAGAGATATCTTGAA	75837					
Qy	221	GlnLysLeuGluArgValTrpGluGlnAsnThrGln-----GlySerIlePro	236					
Db	75838	TATAGATTGGGACAAATCAATCATTTGAATTTGTTATAACTACAAGCTTGTGTAATAATCCC	75897					
Qy	237	TyrValProTrpLeu-----	241					
Db	75898	TACTACATTTGGTGGTTCGCAATAAAGACCACCAACACATCTTCAAGCTAGATTGTGT	75957					
Qy	242	-----ArgAlaProTyrTyrValTyrValTyrIleGlyArg-----	251					
Db	75958	GAACCTTGACCAACATCTCCATACCTCGCACCTATAACACAGCTCCTCTCTAAGCT	76017					
Qy	252	-----Leu	252					
Db	76018	CAGCCATGACCAAGCTATGATTGCAGCACTAGAAAAATTTAAAAAGGTTTACCATAAT	76077					
Qy	253	ProSerValGlyHisAlaLeuHisGluArgVal-----GluArgProPro---	268					
Db	76078	CCTTCTCTGTTCTATCAAAAAGGAGCATTTGGAGAGTGTACTATGATGAAGTGAACCCCTGAG	76137					
Qy	269	-----MetPheProProThrPheLeuTyrThrGlnSerTrpGluAspProGluProAsp	286					
Db	76138	TACAGCCAAATTTAAGAACATTTATATATACGATTCATCGGAAGATTCCTAGAGAAGAT	76197					
Qy	287	MetGluValMetGluIleAsnProLysAspThrValLeuThrLeuThrSerGlyGlyCys	306					
Db	76198	GTTAACATTTCTTAATTAATTAACCAAGAGATACCAATTTTGGCTATTACCACTGCTGGTGAT	76257					
Qy	307	AsnAlaLeuAsn-----LeuLeuValGlnGlyAlaGlyGlnValValSerValAspCys	324					
Db	76258	AATATTTTGCACATGTCCTTTTGCCTAATCCACCGAAGAGAAATTCATGGGGTGATTTA	76317					
Qy	325	AsnProAlaGlnSerAlaLeuLeuLeuLysValAlaIleGlnGlnLeuGluPhe	344					
Db	76318	AATCCATGTCAAGACATTTGACAGAAATTAATAATTAGCAGCATTAGAAGCTTGAGTTTC	76377					
Qy	345	GluAspValTrpGlnLeuPheGlyGlyValHisProArgIleGluGluLeuTyrGlu	364					
Db	76378	ACACAGTTGTGGCAATGTTTGGTGAAGATGATGATGATTAAGATTAACAATATTTTACTG	76437					
Qy	365	LysLysLeuAlaProPheLeuSerGlnThrSerHisAsnPheTrpSerLysArgLeuTrp	384					
Db	76438	AATAAATTTGGCGCATATCTTTTCATCAACGCAATTTCAATACTGG-----	76482					
Qy	385	TyrPheGlnHisGlyLeu-----TyrTyrGlnGlyGly	395					
Db	76483	---TTCAAAATGGAACGAAACAAACATTTGATCTTAATGGAGCCGGGTGTATGACACTGGG	76539					
Qy	396	MetGlyLysLeuCysTrpValLeuGlnCys-----LeuAlaValLeuLeuGlyLeu	412					
Db	76540	TTCAACAAG-----TGGGCATTAAGACTTGGCAAAATGGGTATTCAAAAGTTGCAATCTT	76593					
Qy	413	GlyLysThrValLysArgLeuAlaAsnAlaProThrMetGluGlnGlnArgArgLeuTrp	432					
Db	76594	ACAGATGAAGTTAATATGCTATGCAAGACCTTAAGACCCCTAGAGGAGCAAGATCTATTGG	76653					
Qy	433	AspSerAsnMetLeuIleHisPheValLysAsnGlyProLysProLeuValTrpLeuPhe	452					
Db	76654	GATAAAAAGATA-----AAGCCGGTC-----CTTTTC	76680					
Qy	453	ValLysPheValSerLeuValLeuPhe---AsnLysAlaValLeuTrpPheGlyGlyGly	471					
Db	76681	AACAGATAGTGGTGAATAATCTCGTTGGAAACCCGTTATTTTATGGAGTGCACTTGA	76740					
Qy	472	ValProGlyLysGlnTyrAlaLeuIleLysAlaAspGlyIleProIleGluAsnTyrIle	491					
Db	76741	GTACCACGTAATCA-----GCAAAAATGATGGGCTCTTCAACATTCGAATATATT	76791					

```

QY 492 AlaArgThrMetAspGlyValAlaGluAAsnSerHisValArgLysGlnAsnTyrPheTyr 511
DB 76792 ATTGACACTTTGGACCGCTGTTATTGATAATCTTTCGATTCCGATGATAATTAATTTAT 76851
QY 512 TyrAsnCysLeuThrGlyLysPheLeuArgAspAsnCysProThrTyrThrLeuArgGluAla 531
DB 76852 TATTTGACATTGAAGGGAAGATATTCGAGTAGAAGTTGCTCGATTACCTTAAAGAGGGT 76911
QY 532 AlaPheAlaThrLeuLys-----SerGlyValValAspAsnLeuThrValSer 547
DB 76912 GGATTCAGTCTCTTCTCGTGAAGTCTCGAATCTCTCTAGATAGAGTTAGACTTCAT 76971
QY 548 ThrAsnPhePhe-----MetGluGluLeuLysAlaArgThrTyrThrLysValIle 564
DB 76972 ACTGACACGTTGAAGGATGCTCTGAGCGTTCTCGAAGAAACACAGTTTCGATTGTCTATA 77031
QY 565 LeuMetAspHisValAspTrpLeuAspMetProValAlaAsnGluLeuAlaGlu----- 582
DB 77032 ATTATGGATCACATGGATTGGTTGAT---CCACAAGGAACCTGATGTTGACGGAAGAAAT 77088
QY 583 ---CysLeuAlaLysGlnValAlaProGlyGlyIleValIleTrpArgSerAlaSerLeu 601
DB 77089 CAGGCTTTGGTGGTGGCCCTCAACTCTCGCGTAGGTATATTGAGATCTGCTCTAAA 77148
QY 602 SerProProTyrAlaGluLeuIleGlnLysAlaGlyPheAspValArgCysIleArgArg 621
DB 77149 AGCCCATGGTATATCAAAACTTCGAAAGTTTCGGCTTCAGTTGTTAAAGCAGTAAGTGCA 77208
QY 622 AlaThrGlnGly---TyrMetAspArgValAsnMetTyrSerSerPheTyrMetAlaArg 640
DB 77209 AGATACCTCGGAAATGTATTGATAGAGTTAAACATGATATGCCAGTACCTGGGTTTGCAA 77268
QY 641 Arg 641
DB 77269 AAA 77271

```

RESULT 5

CR382138_20/c

WPCOMMENT

Sequence split into 24 fragments LOCUS CR382138 Accession CR382138

Fragment Name	Begin	End
CR382138_00	1	110000
CR382138_01	100001	210000
CR382138_02	200001	310000
CR382138_03	300001	410000
CR382138_04	400001	510000
CR382138_05	500001	610000
CR382138_06	600001	710000
CR382138_07	700001	810000
CR382138_08	800001	910000
CR382138_09	900001	1010000
CR382138_10	1000001	1110000
CR382138_11	1100001	1210000
CR382138_12	1200001	1310000
CR382138_13	1300001	1410000
CR382138_14	1400001	1510000
CR382138_15	1500001	1610000
CR382138_16	1600001	1710000
CR382138_17	1700001	1810000
CR382138_18	1800001	1910000
CR382138_19	1900001	2010000
CR382138_20	2000001	2110000
CR382138_21	2100001	2210000
CR382138_22	2200001	2310000
CR382138_23	2300001	2336804

Continuation (21 of 24) of CR382138 from base 2000001 (CR382138 Debaryomyces hansenii ch

Alignment Scores:

Pred. No.:	Length:	Matches:
Score:	1,44e-64	110000
Percent Similarity:	864.50	223
Best Local Similarity:	46.1%	105
		Mismatches: 252

Query Match:	25.0%	Indels:	131
DB:	15	Gaps:	20
US-10-620-914-45 (1-648) x CR382138_20 (1-110000)			
QY 41 SerLysLysGlyAspAspHisAlaAlaArgLeuGluSerPheTyrGlyProGlnAlaAla 60			
DB 58985 ACGAAAGAGCCCACTGAGCAACACAGTCGTTAGAACTTTTCTATAAAACCAGGCACAT 58926			
QY 61 AlaPhe-----			62
DB 58925 GTGTATGATAAGCTAGAGAGTATTGTTAAAGGACGTAAGAATGCTTGAGGTTGGCC 58866			
QY 63 AlaAlaArgLeuAlaGluArgSerAsnLeuIleTrpValAspLeuGlyGlyThrGly 82			
DB 58865 ACCGACACATCTCTTAAGAAAGAAAGATCTTGCTGGGTAGACATTCGAGGGGAACATGGC 58806			
QY 83 GluAsnValAspMetMetAlaAspTyrIleAspLeuAlaLys---PheLysSerIleTyr 101			
DB 58805 TCCAATATCGAGTACATGACACGAGGTAATGTGATCTCCAAGAAATTTCAAGGCTGTTAT 58746			
QY 102 ValValAspLeuCysHisSerLeuCysGluValAlaLysLysLysAlaLysAlaLysGly 121			
DB 58745 CTTGTTGATTTATCACCATCGTTTGCGAAGTTGCGAAAAAAGGTCGCTGATAAGAGA 58686			
QY 122 TrpLysAsnValGlnValValGluAlaAspAlaCysGlnPheAlaProProGluGlyThr 141			
DB 58685 TGGAGCAATGTGATGTTCTAGTTGCGATGCTGTGATTTCGGTATTGGGTATGAAAG 58626			
QY 142 AlaThrLeuIleThrPheSerTyrSerLeuThrMetIleProProPheHisAsnValIle 161			
DB 58625 GCGGATTTGGTGACCTTCTCTACTCATTTGTTATGATTTCCAATTTCCATGCTGCT 58566			
QY 162 AspGlnAlaCysSerTyrLeuSerGlnAspGlyLeuValGlyValAlaAspPheTyrVal 181			
DB 58565 GATCATGCTGCTCAAGTTATTAGACAAAACAGGTGTTATCGCCTGTAGACTTCGGTGTG 58506			
QY 182 Ser-----GlyLysTyrAsp-----LeuProLeuArgGln 191			
DB 58505 CAAACCGAGGAGCTTTCGGTTGGTAGAGTAATACCTTAGGAGGTATTGTTAATAGAAC 58446			
QY 192 MetProTrpSerArgArgPheThrArgSerIlePheAspIleAspIleAspIle 211			
DB 58445 ATCCCATGGGCTCTGAGGAACCTCTGCGGTATTTGGTTTGAAGCCCAAGGTTTTTTA 58386			
QY 212 GlyProGluArgArgAlaTyrLeuGluGlnLysLeuGluArgValTrpGluGlnAsnThr 231			
DB 58385 GACCTGTGGAAGACATTTATCTTGAGTACAAATTTGGTACCAATCCCTCAACTGC 58326			
QY 232 GlnGlySerIleProTyrValProTrpLeuArgAlaProTyrTyrValTrpIleGly--- 250			
DB 58325 TACAATAAC-----AAATTGGGACGTAATTCGCTATTTATTTGGGTGGCTGT 58278			
QY 250 -----			250
DB 58277 GACAAATCAAATCGTCATCTTTTATTAAATAGAATCAATTTGTTGGCAACTGAATCCCCA 58218			
QY 251 -----ArgLeuPro---SerValGlyHis 257			
DB 58217 TATCTTGTCTCTTGTGATGATAAGAACAAAGATATCGATATTCTTATTAAGGGTTCAC 58159			
QY 257 -----			257
DB 58157 GAGGCTGCCCTCATGAATTTTCCAAAAAATTTGCTTACCCATCTATCTACTACCAAGA 58098			
QY 258 -----AlaLeuHisGluGluArgValGluArgProProMetPheProProThr 273			
DB 58097 GAAATCTGAGAGTATATCTTCGATGAAATTAGAGATCAATTTTACAGTTTAAAAATCAA 58038			
QY 274 PheLeuTyrThrGlnSerTrpGluAspProGluProAspMetGluValMetGluIleAsn 293			
DB 58037 TATGTGTATGCAATTTTACTTTGGGAAGATCCTCTGAGGATCATTAACATCTTGAAGTTACT 57978			


```
14924..15111))
/gene="B18P24.030"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAE76261.1"
/db_xref="GI:38566960"
/db_xref="UniProt/TREMBL:Q7SDH4"
/translation="MARSAAPRGPGQEAARPTHRFPFPPPAQMKRCGQQQTSET
RNRDFAGPQGNHAPVAGSTRGDDAVLAKDDAPQSGIVSGRCPLGVYFVRPCAGN
ESDMWNEAVRYNPLPIARVKVGLDCMSLVRSPLAVLDSHPCRPPVLAIGFTLRCD
GGSGSRWGTMIILVLRRAAALRAALRMSLVSV"
complement(14161..14342)
/gene="B18P24.030"
/number=1
complement(14343..14411)
/gene="B18P24.030"
/number=1
complement(14412..14477)
/gene="B18P24.030"
/number=2
complement(14478..14565)
/gene="B18P24.030"
/number=2
complement(14566..14717)
/gene="B18P24.030"
/number=3
complement(14718..14923)
/gene="B18P24.030"
/number=3
complement(14924..15111)
/gene="B18P24.030"
/number=4
/notes="29 bp poly (g) tandem repeat"
17595..17623
19883..20827
/gene="B18P24.040"
19883..20827
/gene="B18P24.040"
/notes="similarity to HAPB protein, Aspergillus nidulans,
TREMBL:ENHABP_1
contains EST gb:AA898026, AA898402"
/codon_start=1
/product="related to CCAAT-binding factor HAPB protein"
/protein_id="CAE76262.1"
/db_xref="GI:38566961"
/db_xref="GOA:Q7SGY4"
/db_xref="InterPro:IPR001289"
/db_xref="UniProt/TREMBL:Q7SGY4"
/translation="MMXYAQYQAQNSHNPHMQTYASSAGNSITSPSSOHLPTQTS
PILOTQHQPSPTQGHSMYQYGVQNMHYGMPLQAAALAAATAAGSTNYPNMYA
ADPNLQSPMSGVNSKDVRAQSPSPQMNSTIPQORLSQVSPGVNAQAMLNHA
SRSAVPPMTAAQMPPOSPEMAGVRESPLYVNAKQFHLIKRRVAROKLEQLR
LTNKGKPYLHESRHNHMRPRPGGGRFLTADVAQMERDKVNGDVQDGSSEQSVT
ASKTGTGTRKAESTSGAPNKAAAPSPEDDDASD"
19883..20827
/gene="B18P24.040"
/number=1
21444..21471
/notes="28 bp gcacgcacgcacgcac tandem repeat"
23226..26083
/gene="B18P24.050"
/join(23226..23780,23851..24177,24239..26083)
/notes="no similarity to N-terminal half of the protein
similarity to S-adenosylmethionine:diacylglycerol
3-amino-3-carboxypropyl transferase btaA, Rhodobacter
sphaeroides, TREMBL:AF329857_1
contains EST gb:BF072578, BF072579"
/codon_start=1
/product="related to S-adenosylmethionine:diacylglycerol
3-amino-3-carboxypropyl transferase btaA"
/protein_id="CAE76263.1"
/db_xref="GI:38566962"
```

```
/db_xref="GOA:Q7SGY5"
/db_xref="InterPro:IPR000051"
/db_xref="UniProt/TREMBL:Q7SGY5"
/translation="MGDNSAMASHGGMGNISLPSSVTRTLADLNFTNTTTSKSIFFT
GVAVLVFTTSTNSYRKTKNEDDNEGNPSLSKSLALFCYSCFKPHATAGTCTGTO
QALSFYSQADIIDATRGTLKGRDMLAALASOLRYKVRAGLGLGAGGAGLGRK
QRNGKTCVTAGTGTGTRRKPIWDDVGGGTGMNIEAWKPFVNVSEFFKTVLVDFSPS
LCEVARKFARLGWENVRVICDARKFRLEDYEDVDEGSGSDSPSSGMMGETKP
GHAGAEILTMSYISMPDPFSIIDSLAPHLIIVDFVFAQSKVDFTFRNTYTG
GLMNHVGYFARNFWRSDADRVLSLEPARRDYLYRFGTVLTVNARNNTLGAIPYI
WLGCLKPESTSLPHEIVEHIDALATESPRSSPRLVGKSSATNALAPAVGRTAPE
MRSKAPNTAIENISANLPLSPFFYQNHWRIYYDDQLPKHTQFNDIYIYAFWEDSRV
DRELLNGDDVLAITSAGDNILSYLSPARVHAIDLPNQNHLELUKVASFTIID
YPDWKIFEGEKHPDFRSLLSKLSPLHSGRAFYWLNSAHPFTDPAGRLYDTGSSR
YAIRFRWITLFFCRSAVRLLSTPTLEGQSIYHTKIRPCLLNRFVNLSSDAF
LWASALGPKNOVAMTEADYHRSISSTTPSSKEPSRAEAILHYTTDLPVLSSTSH
LASDNPYLVCVLTGTROCHPDYLSAAHSILSAPGAPDGLRIHTDQEVLARFQP
GTLTVAVYVMSDNDWDFPSPSEKSGRGKARQVRLNRLKVGKGLRSAGVEPWY
VRVVEEGFARRVCGRESGRGDQECIDIRVNMYASCIWLEKMEDLEELVDUSA"
23226..23780
/gene="B18P24.050"
/number=1
23781..23850
/gene="B18P24.050"
/number=1
23851..24177
/gene="B18P24.050"
/number=2
24178..24238
/gene="B18P24.050"
/number=2
24239..26083
/gene="B18P24.050"
/number=3
27765..28499
/gene="B18P24.060"
complement(join(27766..28118,28258..28275,28370..28499))
/gene="B18P24.060"
complement(join(27766..28118,28258..28275,28370..28499))
/gene="B18P24.060"
/notes="match of 21 ESTs
similarity to succinate dehydrogenase precursor (EC
1.3.99.1), YDR178W, Saccharomyces cerevisiae, PIR:A54380
contains EST gb:AI397618, AW718559"
/codon_start=1
/product="related to succinate dehydrogenase precursor"
/protein_id="CAE76264.1"
/db_xref="GI:38566963"
/db_xref="GOA:Q7SGY6"
/db_xref="InterPro:IPR007992"
/db_xref="UniProt/Swiss-Prot:Q7SGY6"
/translation="NASVARSSALLKQVAAQCSVAAGLURVAAFTTSRKSLLPPPPQ
RLEGTVNDPEVPPSPSHGSHYHTFDRVVAAGLPLTVAPFAAGSLNPTMDVLAAT
LIHSHTGEGNIIIVYPSKRVKARKVFTWGLNATVLVLGLALYEFETDVLGTETI
KKVWKA"
complement(27766..28118)
/gene="B18P24.060"
/number=1
complement(28119..28257)
```

Alignment Scores:

Pred. No.:	2,14e-59	Length:	75337
Score:	804.00	Matches:	238
Percent Similarity:	43.2%	Conservative:	112
Best Local Similarity:	29.4%	Mismatches:	256
Query Match:	23.2%	Indels:	205
DB:	15	Gaps:	24

US-10-620-914-45 (1-648) x BX842626 (1-75337)

Qy	19	LeuGLuLyLeuLyLeuSerSerMetLyAspLeuThrValLeuArgHisMet---	37
Db	23674	TTGGAGGAGCTGGGATG---GGCTTGAAGAAAGCAAGGAATGGGAGACATGTGTA	23730

```
QY 38 ---TrpPheGlySerLysLysGlyAspAspHisAlaalaArgLeuGluSerPheTyrGly 56
Db 23731 CCGTGGCCGGGACAGGACCGGACCAAGGA-----GGAAACCCATATGGG 23775
QY 57 ProGlnAlaAlaAlaPheAlaAlaArgLeuAla-----GluArgSerAsn 71
Db 23776 TAGATGTACGCTCTCATCAGAACCCCTTGCCCAATGTTGAAATCCACACACGCTGACC 23835
QY 72 LeuIleTrpValAspLeuGlyGlyThrGlyGluAsnValAspMetMetAlaAspTyr 91
Db 23836 ATCGCTACAAACAGGTGCGTGGGGGACAGGCTGGAATATCAAGCCATGGCCAAAGTTT 23895
QY 92 IleAspLeuAlaLys---PheLysSerIleTyrValValAspLeuGlyHisSerLeuGly 110
Db 23896 GTCAGAGTCTCTGAATCTTCAAGACTGTTTACTAGTGACATTTTCGCGCTCACTTGT 23955
QY 111 GluValAlaLysLysAlaLysAlaLysGlyTrpLysAsnValGlnValAlaGluAla 130
Db 23956 GAAGTGGCTAGGAAGCGGTTTGGCAGGCTGGGGTGGGAGAAATGTGAGAGTTATCTGCACG 24015
QY 131 AspAlaCysGlnPheAla-----ProGluGlyThr----- 141
Db 24016 GATGCTCGCAAGTTTAGGCTTGAGGATTTATAGGATGTTGACGAAGGAGAGTCTGCGCTCT 24075
QY 141 ----- 141
Db 24076 GGAGATTCCTCGCTCTTGTGCGGTTGTTGGGGGAGACGAAGCGGGACACATGCG 24135
QY 142 ---AlaThrLeuIleThrPheSerTyrSerLeuThrMet----- 153
Db 24136 GGAGCTGAGTTGATCACCATGCTCTATAGCTTTTCGATGATGATGATGATGATGATGATG 24195
QY 154 -----Ile-ProPheHisAsnVa 160
Db 24196 GACGCTCTTGGGGTGGTGTGTGCTGACATATCATGACGATGATGACGCGGATTTTCTCGAT 24255
QY 160 lIleAspGlnAlaCysSerTyrLeuSerGlnAspGlyLeuValGlyValAlaAspPheTy 180
Db 24256 TATCGATTCCGCTGAGTCTCTGTGTAGCACCTCAGCGCTTGATTCGCTGCTGAGCTTTTA 24315
QY 180 rValSerGlyLysTyrAsp-----LeuProLeuArgG1 191
Db 24316 CGCCAGTCGAAAGTCGACTTCACATTCGCACTACACAGGCTGTCTTATGAACGACGA 24375
QY 191 nMetProTrpSerArgPhePheTrpArgSerIlePheAspIleAspIleAsp11 211
Db 24376 CGTTGGCTATTTTCGCGCGGAACCTTCTGGCGCTCGTGGTTCGATGCTGACAGGGTGTCT 24435
QY 211 eGlyProGluArgAlaTyrLeuGluGlnLysLeuGluArgValTrp----- 227
Db 24436 TGAGCCAGCTGCTCGAGATTATCTCGAGTACAGGTTCCGGACTGTCTCGACCGTCAACGC 24495
QY 228 -GluGlnAsnThrGlnGlySerIleProTyrValProTrpLeuArgAlaPro----- 244
Db 24496 CCGCAACAACACTTTGGGAGCAATTCCTTACTACATCTGTTGGG-ATGCGCTCAAGAAGC 24554
QY 245 ---TyrTyrVal-----TrpIleGlyAr 251
Db 24555 CCTTTTCTACGTCGAGTCTACCAACAGAAATTGTGAACACATCGATGCTATTTCGCAGAG 24614
QY 251 qLeuPro----- 253
Db 24615 AGTCCCAAGATCATACCCCGCTAGTGGGCAACATTTCTTCTCAGCAACAATCGGC 24674
QY 254 -----SerValGlyHisAlaLeu-----Hi 260
Db 24675 TAGCCTTTGCACTCGCGGCACAGCGCGGAGATGCGCTCAAGAGCCTTCAATACGGCCA 24734
QY 260 sGluGluArgValGluArgPro-ProMet----- 269
Db 24735 TCGAAGACATCTCGGCCAACCTACCTCTCTCCGCTCTTCTTCTACCAAAATCACCACCTGGA 24794
```

```
QY 270 -----PheProProThrPheLeuTyrT 277
Db 24795 GGATCTACTACGACGATCAACTCCCGAAGACACACCCAGTTTCAATGACGAGTACATCTACG 24854
QY 277 hrGlnSerTrpGluAspProGluProAspMetGluValMetGluIleAsnProLysAspT 297
Db 24855 CCTTTACCTGGGAAGACTCGCGCTCGACAGAGAATCTCTTAACCTCGCGGCCGCGACGAG 24914
QY 297 hrValLeuThrLeuThrSerGlyGlyCysAsnAlaLeuAsnLeuValGlnGlyAlaG 317
Db 24915 TGTCTCTAGCCATCACACGCGCGGCGACACATCTTTCTTCTACCTGATGACGAGTCCCG 24974
QY 317 lylGlnValValSerValAspCysAsnProAlaGlnSerAlaLeuLeuGluLeuLysVal 337
Db 24975 CTCGCGTGCACGCCATCGACCTAAACCCAGGCCCAAAACACACCTGCTTGAATCAAAGTCG 25034
QY 337 alAlaIleGlnGlnLeuGluPheGluAspValTrpGlnLeuPheGlyGlyValHisP 357
Db 25035 CCTCTTTTACGACTCTGGATTATCCCGACGCTCTGGAAGATCTTTCGGTGAGGGCAACACC 25094
QY 357 rGArgIleGluGluLeuTyrGluLysLysLeuAlaProPheLeuSerGlnThrSerHisA 377
Db 25095 CCGACTTTCGCTCAGTCTCACTCCAAACTCTCCCTCACCTCTCGCGCGCGCTTTC 25154
QY 377 snPheTrpSerLysArgLeuTrpTyrPheGln-----HisGlyLeuTyrTyrG 393
Db 25155 ATACTGGCTATCCAATGGCACATATTTACCGACCTCGCGGGCGGCTCTATGATA 25214
QY 393 lncGlyMetGlyLysLeuGlyTrpValLeuGlnCysLeuAlaValValLeuGlyLeuG 413
Db 25215 CCGCGCGCTCCCGATACGCTATCCGTTCTTCGCTGGATTTCCACACTCTTCTTCTGCC 25274
QY 413 lylLysThrValLysArgLeuAlaAsnAlaProThrMetGluGlnGlnArgArgLeuTrpA 433
Db 25275 GCTCCCGGCTCGCTGCACTCTCTACTCCACCTCGAAGGGCAACGTTCCATCTACC 25334
QY 433 spSerAsnMetLeuIleHisPheValLysAsnGlyProLysProLeuValTrpLeuPheV 453
Db 25335 ACACCAAGATT-----CGTCCCTGT-----CTGCTCA 25361
QY 453 allyPheValSer---LeuValLeuPheAsnLysAlaValLeuTrpPheGlyGlyGlyV 472
Db 25362 ACGCTTCGTCAACAGCGCTGTCTCAGCTCCGACGCTCTCTCTGCTGGTTCGGTTCGGCG 25421
QY 472 alProGlyLysGlnTyrAlaLeuIleLysAlaAsp----- 483
Db 25422 TGCCCAAGATCAAGTGGCTATGATCGAAGCCGACTACACCCCGTTCTATCTCTCTCT 25481
QY 484 -----GlyIleProIleGluAsnTyrIleA 492
Db 25482 CCACCAACCCAGCAGCAAGAAAAACCAAGCGCGCGCAAGCAATCTCCACTACACAA 25541
QY 492 laArgThrMetAspGlyValAlaGluAsnSerHisValArgLysGlnAsnTyrPheTyrT 512
Db 25542 CTCCACCTCTGATCCCGTTCTCTCCACCTCCACCTTCGCTCGGACACCTTACTTACC 25601
QY 512 yrAsnCysLeuThrGlyLysPheLeuArgAspAsnCysProThrTyrLeuArgGluAlaA 532
Db 25602 TCGCTGTGTCTCGGACATACACAGCCAGTGCCTCCGATTACTTTTCCCTTCGCG 25661
QY 532 laPheAlaThrLeuLysSer---GlyValValAspAsnLeuThrValSerThrAsnPheP 551
Db 25662 CCCACTCTATACTCAGCGCTCTGAGGCTTTGACGCTTACGCTTACGACGAGTAA 25721
QY 551 heMetGluGluLeu-----LysAlaArgThrTyrThrLysValIleLeuMetAspH 568
Db 25722 TACAGAGAGGTGTGGCTAGGTTTTCAGCGGGTACTTTGACAGTAGCGGTGTGATGATA 25781
QY 568 isValAspTrpLeuAspMetPro-----ValAlaA 578
Db 25782 GTATGATGTTGGTTTCGATCCGCTTCGCTCGAGGAGAAAGGAAGGAGGCGCAAGCGCA 25841
QY 578 snGluLeuAlaGluCysLeuAlaLysGlnValAlaProGlyGlyIleValIleTrpArgS 598
```

```

Db      25842 GGAGCAGTGGAGAGGTTGAATCGGCGTGAAGTGGGTGGAAAGGTGTTTGGAGGA 25901
Qy      598  erAlaSerLeuSerProTyTyAlaGluLeuLeuGlnLysAlaGlyPheAspValArgC 618
Db      25902 GC CGCGGAGTGGAGCCGTGTATGTGAGGGTTTGTGGAGGAGGGTTTGGAGCAGGA 25961
Qy      618 ysile-----ArGArgAlaThrGlnGlyTyrMetAspArgValAsnMet 633
Db      25962 GCGTGGGTGTCGTGAATCCGGAAGGGGAGCAGGAGTGTATTGACAGGGTGAATATGT 26021
Qy      633 yrSerSerPheTyMetAlaArgArg 641
Db      26022 ATCTAGTGTGTCGATCTTCGNAAG 26047

RESULT 7
AJ878596
LOCUS      AJ878596      6962 bp      DNA      linear      PLN 03-JUN-2005
DEFINITION Rhizopus stolonifer abcl gene for putative ABC transporter protein.
ACCESSION  AJ878596
VERSION    AJ878596.2 GI:66947996
KEYWORDS  ABC transporter protein; abcl gene.
SOURCE    Rhizopus stolonifer
ORGANISM  Rhizopus stolonifer
            Eukaryota; Fungi; Zygomycota; Zygomycetes; Mucorales; Mucoraceae;
            Rhizopus.
REFERENCE  1
AUTHORS   Cernila, B., Cresnar, B. and Breskvar, K.
TITLE     Molecular cloning and characterization of an ATP-binding cassette
JOURNAL   Unpublished
REFERENCE  2
AUTHORS   Cernila, B.
TITLE     Direct Submission
JOURNAL   Submitted (28-JAN-2005) Cernila B., Institute of Biochemistry,
            Faculty of Medicine, Univ. of Ljubljana, Vrazov trg 2, SI-1000
            Ljubljana, SLOVENIA
REFERENCE  3 (bases 1 to 6962)
AUTHORS   Cernila, B.
TITLE     Direct Submission
JOURNAL   Submitted (03-JUN-2005) Cernila B., Institute of Biochemistry,
            Faculty of Medicine, Univ. of Ljubljana, Vrazov trg 2, SI-1000
            Ljubljana, SLOVENIA
COMMENT   On Jun 3, 2005 this sequence version replaced gi:59889709.
FEATURES             Location/Qualifiers
     source           1..6962
                     /organism="Rhizopus stolonifer"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4846"
     gene             2758..>6962
                     /genes="abcl"
     TATA_signal      2758..2763
                     /genes="abcl"
     mRNA             join(2799..4541,4604..5513,5569..>6962)
                     /genes="abcl"
     exon             2799..4541
                     /genes="abcl"
                     /number=1
     5'UTR            2799..2828
                     /genes="abcl"
     CDS              join(2829..4541,4604..5513,5569..6887)
                     /genes="abcl"
                     /codon_start=1
                     /product="putative ABC transporter protein"
                     /protein_id="CAI47726.2"
                     /db_xref="GI:66947997"
                     /db_xref="GOA:Q5F226"
                     /db_xref="InterPro:IPR001140"
                     /db_xref="InterPro:IPR011527"
                     /db_xref="UniProt/TREMBL:Q5F226"
                     /translation="MKRDHSTKSSVTQSDSGDTOLTNTENHDEKILQEIAAABR
SDVKSSETDKKKKQLSVPIYKIFRFATKLLELLMICIAAIFSAIGAMQPISTIIF

```

```

GQFMTTISTAMASGDYQALVDATHPLVLIPVYMGTAIVLCAYIAQCFMWLTGNSQVRR
INLYXHSIRQMSDFRABESLTRLATQILQIDGIDSKFGLLMVCIGQFLAGF
IIAFVKMFLAVILATIPLLAGTGAAGYFITKTLKSDAYAEAGSVAEQVFSGR
TVYSPQNRFAELYSNRLENAMKGTGIRGQVLGFGGFMFLVLTSTIDRVPEIDPD
TREMLTQSDVLVVFAMLIIGAMALLQLPNLSAVSGCGAAVKIYSTIDRVPEIDPD
SEGLKPSQYNANLEPDMFKYPTPDIITILKLNLTIRPGMTVAFPVSPSGSKSTS
VLIQRDPYDNAGSVFLDHNLRDYNVNLRSQIGVVSQEPVLFNMSIKONLLMGSK
QVNSRVIDPACKANCSFVSQLPDGQYDMVGEHGMGSGGQKQRIAIARAILKNPPI
LLDSEALSALDQSERLVQAALDAASADRTTIVIAHRLSTIRNADLIVMHQGLVEQ
LHSHALLANGVYADLVKQETATKQGVTVTEPDSBELLRERERIEIAOKRAABEL
DEKOTNDHLPRVTTGASSVDVELKERKERKNKAKOSIPMGKVLNOMRSMWLLA
IGVCGNAIAGAVPPCPALLIPARVIALIISPDMSPGPMSTGNIYSFLFVIGCAFIG
FSSQVISFEVAGERITRLRSDFRAFMRQEIQFFDHDNDLGLALSRSLAIDSKNVE
LVTKMGDITQIIVTALIGLSAFSTQWALTILVLCMTPIAFATPYESIKHRGFEK
TKKNEQSGEAGEAIKEIRTVTALNKQNFETKYHRTDHPHLQAKRAYSIGYA
LQOGITLYTNAVAFYAGIRFMANGMIDFQMTCDMAIMLTACQVGRASVFTSTLSKA
KYSATAAPDILEREPEIDPLEGIBPAHQINGDIAFENITTPYPARPDTISFNREN
LHGKSGQTALVCPSCGCKSTTIGMLQRWYDIPISGTVRLDDNNVKYSLNLSHML
VGQSPILFMTIGENISFGVDSIQVQIEQDACSNIHKFVIGLPGQYDTRVGDK
GSQLSGQQRQIAIARAILKRPVLLDDEATLSDSESEKLVOAAIDNILEGGRITTI
TIAHRLSTIQNADLICVKNRGRVIEQGTWELLKLSGVYSDLVYQOQLNAH"
4542..4603
/gene="abcl"
/number=1
4604..5513
/gene="abcl"
/number=2
5514..5568
/gene="abcl"
/number=2
5569..>6962
/gene="abcl"
/number=3
6887..>6962
/gene="abcl"

intron
exon
intron
exon
3'UTR
ORIGIN

Alignment Scores:
Pred. No.:      1,32e-53      Length:      6962
Score:          724.00      Matches:     201
Percent Similarity: 46.2%      Conservative: 118
Best Local Similarity: 29.1%      Mismatches:  224
Query Match:     149          Indels:      149
DB:              15          Gaps:        21

US-10-620-914-45 (1-648) x AJ878596 (1-6962)

Qy      82  GlyGluAsnValAspMetMetAlaAspTyrIleAspLeuAlaLysPheLysSerIleTyr 101
Db      14  GCCTGGAATATTGAATGATGAATGAATATTTTCCATTAGAACAATTTGAAAAGTGATT 73
Qy      102 ValValAspLeuCyHisSerLeuCyGluValAlaLysLysLysAlaLysGly 121
Db      74  TTGTTAGACTTTGACGCCCTTCACTTTGCCAATTCGAAGGAGAGATTGCTAAGAGAGGA 133
Qy      122 TrpLysAsnValGlnValValGluAlaAspAlaCysGlnPheAlaPro----- 138
Db      134 TTTAAAAATGTCTGTCTGTGTTGTCAAGAAGCGGCTAATTTTCAGATAGGATAGACTA 193
Qy      139 GluGlyThrAlaThrLeuIleThrPheSerTyrSerLeu----- 151
Db      194 GAAGGCGGTGAGGGTTAATCACTGTTCTCTATGC-ATGTAAGTTTAAGAATAGTAGGAT 252
Qy      152 -----ThrMetIleProPro 156
Db      253 GTGTATGTGTGTATGTATGTGTAAAGTATGTCTATTAGTGTCTATAGGTGATCAT 312
Qy      157 PheHisAsnValIleAspGlnAlaCysSerTyrLeuSerGlnAspGlyLeuValGlyVal 176
Db      313 TTTATCCAGTGTATGATGCTATTCAGTCTCTTTTGGCACCAGAGGTATTATTGGAGTG 372
Qy      177 AlaAspPheTyrValSerGlyLysTyrAspLeuProLeu----- 189

```



```
QY 40 GlySerLysLysGlyAspAspHisAlaAlaArgLeuGluSerPheTyrGlyProGlnAla 59
Db 26888 GGTAGTGCAGACCAAGAACCCAC-----CTCGAGCGTCTTTACGCGCGGCAAGCC 26938
QY 60 AlaAlaPhe----- 62
Db 26939 GATTATATACAGACAGCACGCTCCAGCTCTCAAGGGTAGAGAGACCATGCTTCAGCTG 26998
QY 63 ---AlaAlaArgLeu-----AlaGluArgSer 70
Db 26999 TTGGCTGCCCATCTCAAAAGCTCAGCCTATGCTGAGGCTTCCCGTTAGTGCACCCCTCCAAA 27058
QY 71 AsnLeuIleTyrValAspLeuGlyGlyThrGlyGluAsnValAspMetMet----- 88
Db 27059 CCGAGGATATGGGTGGATTGGGAGCGGTACCGGTTGGAAACATTGAAAGATGTACGTC 27118
QY 88 ----- 88
Db 27119 TTTTGGCTCTGTACAGTCAAAACGTTTTGAAACTGTGACTAAACCTGAAACCTGTG 27178
QY 89 -----Ala-AspTyrIleAspLeuAlaLysPheLysSerIleTyrValValAspLeuCy 106
Db 27179 CCATAGGAGTATGATCTCTCCATTGACCTACTTTGATGCGATATACATCATCGACCTATG 27238
QY 106 sHisSerLeuCyGluValAlaLysLysLysAlaLysAlaLysGlyTyrLysAsnValGI 126
Db 27239 TGAACCCCTCTTGAAGTGGCCAGGCGCCCAAGATCAAGGCAAGGGGTGGAAGAATGTCCA 27298
QY 126 nValValGluAlaAspAlaCyGlnPheAlaProGlu----- 139
Db 27299 CGTCTGTGCAAGATGCTAGTAGGTTGTTTGCAGAGTGGGAGCGGGCAGTGGA 27358
QY 140 -----GlyThrAlaThrLeuIleThrPheSerTyrSerLeuThrMet----- 153
Db 27359 TCCGAGGGGAAGTCTGAGCGCGATCACCATGAGCTATTCGCTCTCCATGGCAAGTTTCA 27418
QY 154 -----IleProProPheH 158
Db 27419 GACATTTCCGACCTTTTCTCCCGATGAACCTGATTGTTCTCACAAGATACCTCCCTTTT 27478
QY 158 isAsnValIleAspGlnAlaCySerTyrLeu---SerGlnAspGlyLeuValGlyVala 177
Db 27479 ACCAAGTCTTGGATAGATGATCAAGTCTTGATCTCAAGGGGATGATGGCGGTG 27538
QY 177 laAspPheTyrValSerGlyLysTyr-----AspLeuProLeu- 189
Db 27539 TCGACTTTTACACTCTCGGAGAGTAGGTAAACAAGGAACGCGTAAGCCCTCTCCCTTTT 27598
QY 189 ----- 189
Db 27599 TTTTGAATATCACTATCAGATGCTCAGCTAATAACTTTCTTTATACATAGGCAATCGGTA 27658
QY 190 -----ArgGlnMetProTyrSerArgArgPheTyrPheTyrPheAspIleA 207
Db 27659 CAGCGAGCAACGCTGATCTTGTTCTCAAAATGGTTCTGGAGTGTGCTGCTGACCTCG 27718
QY 207 spAsnIleAspIleGlyProGluArgAlaTyrLeuGluGlnLysLeuGluArgValT 227
Db 27719 ATGCGGTTCATCTTCACGGGTCTCGACGAGTACCTGGGAATACAAAATGGGCACCATCA 27778
QY 227 rpGluGlnAsnThrGlnGlySer-----I 235
Db 27779 AGACGTACATAGCGGGAACAACATTTTGAACACTTGGTTTCATACAAAATCCCTAGTATG 27838
QY 235 leProTyrValPro---TyrLeuArgAlaProTyrTyrVal----- 247
Db 27839 TCCCTATAAACCTTTTGGCTTGGAGGGGAAAATTTGGATAGATTGACATATTAATATCAG 27898
QY 248 -----TyrIleGlyArgLeuProSerValGlyHisAlaLeuHisGluGluArgV 264
Db 27899 CTATGCTCTCTCGGTT-----GCTCAGCACAACGATGC 27934
QY 264 alGluArgProPro----- 268
```

```
Db 27935 TTCTGGCTCGCCAAATCGTTTACTTTTGAAGCGGTAATCGTCTCGGCCAAAGCGACCT 27994
QY 269 -----MetPheProProThrPheLeuTyrThrGln----- 278
Db 27995 CGGCTCTCCTCACCCCTACCTCGCTTTTACCAACTCTCCATCTGTATTTGGCTCTCCTTC 28054
QY 278 ----- 278
Db 28055 ACCATGCTGGAAATGCCAGATTAGTTCTTAGGACCGAGTGGCGGCACAGCGACTCAAAC 28114
QY 278 ----- 278
Db 28115 AATTTTGAAGTGGTGGTCTCTCTTAAGCCGTTTCTACTACCTCTGAGAAAGTGGGTAT 28174
QY 278 ----- 278
Db 28175 TTTTGTGGTCTCTGTTGGATGGGAAAACTCATGAGACACAGCGCATGGAGGATACCT 28234
QY 278 ----- 278
Db 28235 TACCTTGAAGAAAAAATCCACGAGCAATTCAGGACTCACATTTACGATGGGTGAGTCA 28294
QY 278 ----- 278
Db 28295 GCGCCGGGGGGGGGTTCTTGATCCCAAAATACACACACCCGAAACTAACTTTTTCGA 28354
QY 279 -----SerTyrGluAspProGluProAspMetGluValMetGluIleA 293
Db 28355 TGTGTTGACTCAATAGACTTGGGAAGATCCGGCGGTGACGTCACAGAAATTCGAACATCA 28414
QY 293 snProLysAspThrValLeuThrLeuThrSerGlyGlyCysAsnAlaLeuAsnLeu--- 311
Db 28415 ACAAAGATGACACATCTCTTGCATCCTCTGCAAGTACATGTGCTCCA-TTACGCC 28473
QY 311 ----- 311
Db 28474 TTGACGGCCAAACCTCGAGGATTCACGCGCTGCTGAGTATTATTCATTAATAATCTCAT 28533
QY 312 -----LeuValGlnGlyAlaGlyGln-----ValValSerValAspCysA 325
Db 28534 CAATCTTTTACATCATCCAAAACCTTCACCAAGATGAATCTTTACGATGATAGATATGA 28593
QY 325 snProAlaGlnSerAlaLeuLeuLeuLysLysValAlaIleGlnGlnLeuGluPheG 345
Db 28594 ACCCGTGTCAAGTGCATTTTAGAGTCAAACTTCAGCTATCCAGGCAATTCGAATACA 28653
QY 345 luAspValTyrGlnLeuPheGlyGlyValHisProArgIleGluGluLeuTyrGluL 365
Db 28654 ATGATTTCTGGCTCATTTTGGTGAAGCGCCGCCACAGAGTTTGAACAGCTTTTGACTA 28713
QY 365 yslLysLeuAlaProPheLeuSer----- 372
Db 28714 CCAAGCTATCTCCCTCTTGTCTTCGTAAGTCCCGCCCTCTAAATAAAGACTACAT 28773
QY 373 -----GlnThr-SerHisAsnPheTyrSerLysArg 382
Db 28774 CTTTCTCTAACATGGTTGGGGGTTTTACAAGACACGCGCTACGCTTATTGGAAAACCCAT 28833
QY 383 LeuTyrTyrPheGlnHisGlyLeuTyrGlnGlyMetGlyLysLeuCyTsrVal 402
Db 28834 ACGTCTCAGTTCTCAGCAATTTCTACTTTTCGAGGCTACAGCGGC-----TGGGCT 28884
QY 403 LeuGlnCysLeuAlaVal-----ValLeuGlyLeuGlyLysThrValLysArgLeu 419
Db 28885 CTCGCTCGCCAGATTCATCTTTATCGTGGTGTAGAGGGGATGTCAAGCGCTG 28944
QY 420 AlaAsnAlaProThrMetGluGlnArgArgLeuTyrAspSerAsn----- 435
Db 28945 TGTGAGCGGACTTCGAGCGGTGAACAGGAAGGATTTGGCAAAAGAAAGATTAGACCCGTT 29004
QY 436 -----MetLeuIleHisPheValLysAsnGlyProLysProLeuValTyr 450
```



```

CDS
complement(3..182)
/locus_tag="RB2"
/codon_start=1
/transl_table=11
/product="hypothetical protein-signal peptide and
transmembrane prediction"
/protein_id="CAD71330.1"
/db_xref="GI:32442812"
/translations="MRLPFSALFAFSVFTVCGCGGNGEVIEDTRSPAEOEQED
YKQKEMDSATNS"
39..260
/locus_tag="RB3"
39..260
/locus_tag="RB3"
/codon_start=1
/transl_table=11
/product="hypothetical protein-signal peptide and
transmembrane prediction"
/protein_id="CAD71331.1"
/db_xref="GI:32442813"
/translations="MLFVIFLLHFGWRAGVDFHFVAAASATNSDRNAERKQGGK
QSHRYCSMDARNECEKLVRELLILRISP"
complement(399..1514)
/locus_tag="RB5"
complement(399..1514)
/locus_tag="RB5"
/codon_start=1
/transl_table=11
/product="hypothetical protein-signal peptide and
transmembrane prediction"
/protein_id="CAD71332.1"
/db_xref="GI:32442814"
/translations="MTKRLPTRRRLVLAAGLITPLAAASTAKAGFGEWETGSPY
YEDAWYVSEWEDGNPTDEAIGKWDNETYDRDEALTSSDNDIDWTKNDHGTY
ADYDSSTFYDYHGYSDYDSDNGNYEYTSNYDYDNDGVDYDAFASYDTDDGG
IYDDATYIFSNASQNSQKTAQKADQONSAKSQSFTGTTTRAKVKVTPATNVV
VELQNSQEGSMIADLPADQDLPQDLNDTLTVKGSFPKAGDTSVLLSRSEIHNGQ
MOVERSREYSGTIESTKTVDRGQORQAKIKTDGKKMLVDLGRKQDLDTNITEGS
QVTINGPAQVQDRVMLLATKVILDGDTIEISMAQ"
1548..1706
/locus_tag="RB7"
1548..1706
/locus_tag="RB7"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAD71333.1"
/db_xref="GI:32442815"
/translations="MRSRSQSVVQETNGRKVDKDKDGAARVAQPNRFGKLSGVAAQ
RGSKQKVQ"
complement(1733..5263)
/locus_tag="RB8"
complement(1733..5263)
/locus_tag="RB8"
/function="DNA synthesis and replication; transcription"
/notes="best DB hits: BLAST: pir:D83545; probable helicase
PA0799 [Imported] - Pseudomonas; E=7e-95 embi:CAA67095.1;
(X98455) SNF [Bacillus cereus]; E=8e-94 pir:C72027; swianf
family helicase 2 - Chlamydomophila pneumoniae; E=4e-87 COG:
PA0799; COG0553 Superfamily II DNA/RNA helicases. SNF2
family; E=7e-96 PFAM: PF02803; Thiolase, C-terminal
domain; E=0.11 PF00176; SNF2 and others N-terminal domain;
E=3.9e-64 PF00271; Helicase conserved C-terminal doma;
E=1.2e-24"
/transl_table=11
/product="helicase"
/protein_id="CAD71334.1"
/db_xref="GI:32442816"

/db_xref="GOA:Q7UZE8"
/db_xref="InterPro:IPR000330"
/db_xref="InterPro:IPR001650"
/db_xref="InterPro:IPR011545"
/db_xref="UniProt/TREMBL:Q7UZE8"
/translations="MDWIAEDQLGSLGLLMGEAVERLVDSHDQGFDSRAAMLTKPTT
LTQDNGVRFDFQVGNREKQQTVALIEIVQDEDDPGVGELSSCCGCVAFLRDN
SCPHTLAAAWLQEQIARRNGQLIEFGLGKADAAVAGKQFVDFLLRRKNSRPAP
AEDABETPLQWRVVRNMAQAPLITTFEQKRGKGKGTGREVSFDDLLNLT
DPLDSKIASLTSPVTHQDNNVAIPEALRMLAGHENVAWQDONALPVSFVCSQLLTT
LQPEVBSQKPKASKSOTDGRASKKTAARTKSSGKSGKSADELGNVTATWTM
PSEILDSIAPSDAGLIPDDLIIPDDEKVQFQPLSDVDGIDIDVQCEIVVGLSPV
EPAILADADEDRIVLCTLADPAAGIIOFLRGKQVTRULDHAHATSRALAAAGSDS
LIRVLPQLAGIIPAPCELVQMRPFAGLKVSRVYESRLRSLIPGDAPELVP
CLAETGPIRLQDLKAELEAEKIIVQFGLSRLSHESHTMEWALTDDHALDLRLHD
AGDESPIRVMEGETFVRGEITPKSLRIKIDSDPWFQMTGTVTVGVDPVLDLLM
AVREDROLVRVDRRFPAKISEAPKRLTOLGDTLVSEKQLRAGAAVPLARVLDGD
VTELTAMHQAIERLSDLSQMSDORPEGLDATLDYQDGYRWLARLSRWVGGVLA
DMGLQKNTVTLGILVERGVPALVVAFTSVGENVWREAEKFTPALTKLYRDCDRD
EVIKTAGPNLLIVSYQLQDAKFPASGHTLVLDEAOFKNSQTKTSQAIRITQIA
DWRLGLSGTLENHIGELMSLFTLSGLLGSQWRFSRFAEPIERHSADRRESLAR
LVRFVLRRTKDKVLKELPRTETILRAELSAPEKLYEDARLAALAEITAGGAPSH
EGRRRIQTLISMLTRQLACHPSLVEPSMKGTSKLLQLFLSLVEELREGDHRALVFSQ
FYKHLSVRAALDERGISYQYLDGATPSHERQRRVDAFQNGEGDLFLISLKAGGTGLN
LTAADYVLHLDPPWMPAVEDQADRAHRIGQERAVTVYRLVAERTIEEQILQLHADKR
ELVAGILDGTDAAKLQTDLLIKAENG"
complement(5458..5850)
/locus_tag="RB13"
complement(5458..5850)
/locus_tag="RB13"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAD71335.1"
/db_xref="GI:32442817"
/db_xref="UniProt/TREMBL:Q7UZE7"
/translations="MHLPGNORFKDKTKNYRETSETLRTILATETVLTVPARHPF
PRRSELMEETPESDPQLSKCPKIPSGTIPGNKVSCSPVIPKELPHITTCBPPLSNR
AGNRVDELKRGVGVGFWGTRPVLRLSP"
5836..7386
/locus_tag="RB14"
5836..7386
/locus_tag="RB14"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAD71336.1"
/db_xref="GI:32442818"
/db_xref="UniProt/TREMBL:Q7UZE6"

Alignment Scores:
Pred. No.: 5,31e-28 Length: 311050
Score: 460.00 Matches: 139
Percent Similarity: 43.4% Conservative: 96
Best Local Similarity: 25.7% Mismatches: 210
Query Match: 13.3% Indels: 96
DB: 1 Gaps: 18

US-10-620-914-45 (1-648) x BX294133 (1-311050)
QY 140 GlyThrAlaThrLeuLeuThrPheSerTyrSerLeuThrMetIleProPheHisAsn 159
Db 181119 GGTACTGCAACACGGGTG-----ATTGGTGTGAGAACTGCACCGGACTGGTGAAC 181069
QY 160 ValIleAspGlnAlaCysSerTyrLeuSerGlnAspGlyLeuValGlyValAlaAapPhe 179
Db 181068 GTCATGACGGTGCAGTGCATTGG-----TCGACGCTACGACGAGGACTTATTCCTTC 181015
QY 180 TyrValSerGlyLys-----TyrAspLeuProLeuArgGlnMetPro 193
Db 181014 GTCTGCCACACAGAGAAACGCTGCGTGGTGTGTTTCATGCCACAAACGCCAAGTA 180955
QY 194 TrpSerArgArgPhePheTrpArgSerIlePheAspIleAspAsnIleAspIleGlyPro 213

```



```

Db      1345 CCCTCGCGGAGAACTACTTCGCTTCCAGGCCATCGCGCGCTACTCGCGCGCCGCGC 1404
Qy      524 -----CysProThrTyrLeuArgGluAlaAAlaPheAlaThrLeuLysSerGlyVal 540
Db      1405 GAGGGCGCGCTGCGCGCTTATCTCGAACCCACCGCTTCGAGACGCTGCGCGAGAAC--- 1461
Qy      541 ValAspAsnLeuThrValSerThrAsnPheMetGluGluLeuLysAlaArgThrTyr 560
Db      1462 GCGGGCGGGTGAGATCGAGATCGAGAACCGCAGCTGACCGAGCGCTCGCGCCGAGACCGGAG 1521
Qy      561 ThrLysVal-----IleLeuMetAspHisValAspThrPheLeuAspMetProValAla 577
Db      1522 GAGAGCATCCACGGCTTCACCTGCTCGATCGCGCAGACTGGATGACGGAGCGCGAGCTG 1581
Qy      578 AsnGluLeuAlaGluCysLeuAlaLysGlnValAlaProGlyGlyIleValIleThrArg 597
Db      1582 ACCGCGCTCTGGCGGCGAGTGCAGCGCACTGCAGCGCGCGCGCGCGGCGGTGATCTTCGCG 1641
Qy      598 -----SerAlaSerLeuSerProProTyrAlaGluLeuIleGlnLysAlaGlyPhe 614
Db      1642 ACCGCGCGGGCGCGCACTGCTGCCC-----GGCGGAGTGCCCGAGGAGATCTCGGG 1695
Qy      615 AspValArgCysIleArgAlaThrGlnGly-----TyrMetAspArgValAsnMet 632
Db      1696 CACTGGCGCGCCGACCGCGCGCGGCGGACAGCGCGGCATGCGCGCGCGGTTCGCGCATC 1755
Qy      633 TyrSerSerPheTyrMetAlaArgArgLysGlyAla 644
Db      1756 TACGGCGGCTTCCACCTCTACCGCGCGGAGGAGCGCA 1791

RESULT 12
AR551999
LOCUS      AR551999      933 bp      DNA      linear      PAT 08-OCT-2004
DEFINITION Sequence 7130 from patent US 6747137.
ACCESSION AR551999
VERSION   AR551999.1 GI:53945174
KEYWORDS
SOURCE    Unknown.
ORGANISM  Unclassified.
REFERENCE 1 (bases 1 to 933)
AUTHORS  Weinstein,K.G. and Bush,D.
TITLE     Nucleic acid sequences relating to Candida albicans for diagnostics
          and therapeutics
JOURNAL   Patent: US 6747137-A 7130 08-JUN-2004;
          Genome Therapeutics Corporation; Waltham, MA
FEATURES  Location/Qualifiers
           source          1..933
           /organism="unknown"
           /mol_type="genomic DNA"

ORIGIN

Alignment Scores:
Pred. No.:      1.25e-21      Length:      933
Score:          352.00      Matches:      95
Percent Similarity: 40.5%      Conservative: 35
Best Local Similarity: 29.6%      Mismatches: 97
Query Match:    10.2%      Indels:      94
DB:             6           Gaps:         8

US-10-620-914-45 (1-648) x AR551999 (1-933)

Qy      144 LeuIleThrPheSerTyrSerLeuThrMetIleProPheHisAsnValIleAspGln 163
Db      7 TTGATTACTTTTTCGATTTCATTCGATGATCCAACTTTCATGCTGCTATCGATAAT 66
Qy      164 AlaCysSerTyrLeuSerGlnAspGlyLeuValGlyValAlaAspPheTyrValSer--- 182
Db      67 GCTGTTTCTAAATAGATATGAGAGGTATATTGCCACTGTAGATTTTGGTATTCAAGC 126
Qy      183 -----GlyLysTyrAsp-----LeuProLeuArgGlnMetPro 193

```

```

Db      127 AGTGACACCTCAATGGGTGCTATTATACTGTTGGGGTGGTTAACAGGGACATTCCT 186
Qy      194 TrpSerArgArgPhePheThrArgSerIlePheAspIleAspAsnIleAspIleGlyPro 213
Db      187 TGATATTACGTAATTTTGGAGATTTGGTTGAAGCTGATAAAGTGTTTGGATTCT 246
Qy      214 GluArgArgAlaTyrLeuGluGlnLysLeuGluArgValTrpGluGlnAsnThrGln--- 232
Db      247 TCAAGAAGAAACTATTTTGGAATATAAATTTGGTACCGTCAATCTTTGAATTCATCAAC 306
Qy      233 -----GlySerIleProTyrValProTyrLeuArgAlaProTyrTyrValTrpIle 249
Db      307 AAGGCTTTGGGTAAATC-----CCCTATTATATTGGATT 342
Qy      250 GlyArgLeuProSerValGlyHisAlaLeuHisGluGlu-----ArgValGlu 265
Db      343 GGTGTGATAAATCAAAATCACACCATTTTAGAAAGATTGAATTGTTAGCCACTGAA 402
Qy      266 ArgProProMetPheProProThr----- 273
Db      403 TCCCTTACCTTGTCTCCAACTACAACCTCCAATCGCTAATCAACTTGAAGATATTCCAAT 462
Qy      273 ----- 273
Db      463 TCTAAAGTCAATGAAGTCTGTTTAATCAACTTGCACAAAAAATTTACCTTACCCATCAATG 522
Qy      273 ----- 273
Db      523 TACTATCAAAAGGAATATTGGAGAGTCTACTATGATGAATGAATCCATTGTATGAACAA 582
Qy      274 -----PheLeuTyrThrGlnSerTrpGluAspProGluProAspMetGluVal 289
Db      583 TTTAAAAACCAATACATTTATGCTTTCACCTGGGAAGATCCTCGTGAAGATCATAAACTT 642
Qy      290 MetGluIleAsnProLysAspThrValLeuThrLeuThrSerGlyGlyCysAsnAlaLeu 309
Db      643 TTGAATTTTACCAGTGTGATGATCTGTTTGGCTATTACTTCACTGCTGTGATATAATTTTG 702
Qy      310 AsnLeu-----LeuValGlnGlyAlaGlyValValSerValAspCysAsnProAla 327
Db      703 AGTTATGCTAGTTTACCAACACACCAACCAAAAGATTCATGCTGTGTATCTTAATCCATGT 762
Qy      328 GlnSerAlaLeuLeuGluLeuLysLysValAlaIleGlnGlnLeuGluPheGluAspVal 347
Db      763 CAAACCATTTATTAGAAATTTGAATTTGGCTAGTTTGTAGATGCTCTTCTCAAGAACAAAT 822
Qy      348 TrpGlnLeuPheGlyGluGlyValHisProArgIleGluLeuLeuTyrGluLysLysLeu 367
Db      823 TGGTCAATGTTTGGTGAAGGTAAATTCGAAAAATTTCAATGATCTTTTGATTGATCTTTG 882
Qy      368 AlaProPheLeuSerGlnThrSerHisAsnPheTrpSerLysArgLeuTyrTyrPheGln 387
Db      883 GCGCCGCACATGCTTCTAATGCG-----TTCAA 912
Qy      388 His 388
Db      913 CAC 915

RESULT 13
AR5008128
LOCUS      AR5008128      10129 bp      DNA      linear      BCT 28-MAY-2004
DEFINITION Agrobacterium tumefaciens str. C58 circular chromosome, section 186
          of 254 of the complete sequence.
ACCESSION AR5008128 AR5007869
VERSION   AR5008128.1 GI:15157252
KEYWORDS
SOURCE    Agrobacterium tumefaciens str. C58
ORGANISM  Agrobacterium tumefaciens str. C58
          Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
          Rhizobiaceae; Agrobacterium.
REFERENCE 1 (bases 1 to 10129)
AUTHORS  Hankle,G., Slater,S.C. and Goodner,B.
TITLE     Complete Genome Sequence of Agrobacterium tumefaciens C58

```

(Rhizobium radiobacter C58), the Causative Agent of Crown Gall
Disease in Plants
Unpublished
2 (bases 1 to 10129)
Hinkle, G., Slater, S.C. and Goodner, B.
Direct Submission
Submitted (14-AUG-2001) Cereon Genomics, 45 Sidney Street,
Cambridge, MA 02139, USA
FEATURES
Location/Qualifiers
1..10129
/organism="Agrobacterium tumefaciens str. C58"
/mol_type="genomic DNA"
/strain="C58"
/isolate="Cereon"
/db_xref="taxon:176299"
99..1175
/gene="AGR_C_3839"
99..1175
/gene="AGR_C_3839"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_C_3839p"
/protein_id="AAK87865.1"
/db_xref="GI:15157253"
/translation="MNSGKFRGRSALLSSVTGLALASACTSVBYAAKEGPNPKI
ATVTSQAQTSQPTATAEAVAGTAAVAPVADAAASQAALPTAVAVKGRVAL
PPASEIAAAAIATOMQPTAEVAGAAATTAALVPTHTTAAAIETAAIAAK
PAMHNAADLPVAAVQKGFPPAPPPGSPSIGTELTAERKVIPLPKPSTVLAYAG
PTNAALAIKONESMTAPLRTAPAGREKLSGLTKYATWYDVPEDLVHRVQRESYN
PGAYRNGFLMIRIATARSALGDPGASGLFPAETNLKYAVLYLKGAVVAGNDRDN
ANLYARGYYIDAKRGMLHLRN"
1517..2767
/gene="AGR_C_3843"
1517..2767
/gene="AGR_C_3843"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_C_3843p"
/protein_id="AAK87866.1"
/db_xref="GI:15157254"
/translation="MTSAAPKTFGSKTKLKSALLQKALSGLSERFGVLFGLV
YPOIWDPIDEMBELGHEGRIVITIGSCNNLAYSRLNPASIDVVDLNPPIALNK
LKLAAPHLPAHQVDVYHFGRACTNSVGYDFIAEHLDTATKAYWSKRTLSGRRI
SVFDRNIYRTGLIGRIGAGHIMARLHGKLTETAKRTLDEORQFDSKVLPLDKP
VVRWLTNRKSSLLFGIGIPROVDELASLSDGCTVASLKERLEKACNFPPLSDNYEAW
QAFARYPEPHEGALPAYLKPEYKEIKRNTARVAVHHATYTELKSRKPAANGVRYIL
LDAQDWTVDQLNELWSQISRTAASGARVIFRTAEKSVIEGRSLSPDIRNQWVYLEER
SNELNADRSAYGGPHIYQAWA"
2764..3483
/gene="AGR_C_3844"
2764..3483
/gene="AGR_C_3844"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_C_3844p"
/protein_id="AAK87867.1"
/db_xref="GI:15157255"
/translation="MKTIGENVGLADSAHAGLMDRMRYRQRIHYDITRKYLLGRDR
ISGLDVPKGTLLIEICGTRNLLASRFPDAKFGLDISAEMLLTASENFAKAE
FILRVADATFASRSECGPDGRVNIPIYALSMIPWEKAIIEOALAAKPKGSLHVD
FQOQOLPKWFTLLQAWLTRFHVTPRANLRYVLANMAGRFDGSLVFEIARGYAWRA
VITLPVAEAPQPKIHELLADA"
3434..4498
/gene="AGR_C_3846"
3434..4498
/gene="AGR_C_3846"
/note="HYPOTHETICAL 32.0 KDA PROTEIN IN FBAB-THID
INTERGENIC REGION"
/codon_start=1
/transl_table=11

/product="AGR_C_3846p"
/protein_id="AAK87868.1"
/db_xref="GI:15157256"
/translation="MPKPRSRSTAYVLTDPDSRIKKPDRPHAASSRLSLSGPISALK
POLAWNPAICEVQKTSPLPYQFVHVMBKSGDTABRFAAGWATLIIIRVMBLRPFA
LLSALLLVSGCTSTSYDLLETASVSRPKFSDTDQDFGVNNPHRHEVHGIDVSKWGD
VDMQTVKSGSVFVFKATEGSDRIDPKGDFHWRSAASANILHAPHYFYFCSADAQ
ADMFIISNPKAEVTLFPVLVDVWNPSPCTKTRPAFGIVRAEORFLDRLEAHVYGRKP
IITYSDVFRHDLNLVQGFQYHFWRSVAAHPAKIYEDRWKFAFWQYATATGVVPGVNGPT
DINVPAGSEKQWRKWTASAK"
4622..5848
/gene="AGR_C_3849"
4622..5848
/gene="AGR_C_3849"
/note="hypothetical protein PA3992 {imported} -
Pseudomonas aeruginosa (strain PAO1)"
/codon_start=1
/transl_table=11
/product="AGR_C_3849p"
/protein_id="AAK87869.1"
/db_xref="GI:15157257"
/translation="MPTFRSRSTVFAAALLSAGSALATPPVASPDDPKQVACGGDL
GAFLEGVKAEEAVAKGVPAALIEKALAGAOIDPKVLSRDRAGVFRGTFLFESORTVSO
ARLDIGRKKLOELSSSTFAEAEFEGVPAGVIAAPFAWMETDFGAVQSGDFTRNALVTL
HDCRRPELPQQLALIEWQHGDLDPATNGAWRIGQVQMLPKDITAFGVDDGN
GHINVDKSDAPDAILTAAKFIQHLGFKKGPPMOEVSPELNPWKSGLSGLTAGDFW
ALGVTPRDGNKFPASLPAALIMPOGRKGPFTFYPNFNIYLEWQSFITYTTSAYFAN
RLMGAVDQVKNPDPQGFDDATMKELQTKLEARGHDVKGVDGILGSGTRVAIOKEQOQL
GMPADGMPYQALLQAL"
complement(6313..7251)
/gene="AGR_C_3850"
complement(6313..7251)
/gene="AGR_C_3850"
/note="5.10-methylenetetrahydrofolate reductase (FADH2)
(BC 1.7.99.5) - Salmonella typhimurium"
/codon_start=1
/transl_table=11
/product="AGR_C_3850p"
/protein_id="AAK87870.1"
/db_xref="GI:15157258"
/translation="MWCKRRFYFAPGGHMLBSETARRDDIELSFEFFPPKNAEMDA
QWETVAELKWDPPFVSVTYAGGSKTAPTLDAVRMIGDSLHPHTAHLTCVDASRD
DVHQVIEEPNAGVVRHVALRGDPSTGIGTAYOPHPEGYENAAALVKGURBIGDFEIS
VSAYPEKHPEASSETVIDMLRKADNGDTRALTOFFPDNDVFERYMERTAAAGITIP
VVPGMPIQNLTKLRFACRCGTSVPRFLDERPEGDEDELEGRARVAEVAABQIADL
ORRGNEPHLYTMNRAPLVDAVLENLGLQLRRSAA"
complement(7208..8221)
/gene="AGR_C_3852"
complement(7208..8221)
/gene="AGR_C_3852"
/note="hypothetical protein XF1617 {imported} - Xylella
fastidiosa (strain 9a5c)"
/codon_start=1
/transl_table=11
/product="AGR_C_3852p"
/protein_id="AAK87871.1"
/db_xref="GI:15157259"
/translation="MAFGDLKLVLLKAAGETPRFELLALLSAGDITVTDLTILGOS
QPRISRHLKLTTEGLIERYQEGANAYFLRQDGEPAVVVYRQLLGAASPSDAVLQRDS
EKLATVQRAERAQDIYFSRNASAMDELRLRHVSDAEVAGALLKVGTFITPDSLLDGL
TGTGRLIKVLSGIYRAIGVADSRDLMSVARANLKGAGVNASVRHGDILNPLEQD
FLIIVHQLVHFLDQPEALSEARMLRPGRLIVIDLAPHSPFYLRDDEHVAHVRLGFS
HPLIEMWKKKAGLSLERAVDLHSEKASDEALDVTIWARDQRLLMADDVQEAVALLS
GRA"
complement(8383..10032)
/gene="AGR_C_3855"
complement(8383..10032)
/gene="AGR_C_3855"
/note="(AB041227) ABC transporter"
/codon_start=1
/transl_table=11
/product="AGR_C_3855p"
/protein_id="AAK87872.1"

```

/db xref="GI:15157260"
/translation="MARQRTYHMGALNKSYGAKKLVNHLSPVDPKIGILGPNAG
KSTVLKIMAGLDEYGEAWLASGATLILEQEPKLDENKVTNMEGVASKTAIVD
RYNELMNYSDETADSGAKQDMIDSONLMDQNVEMAMDLKCPFGSDAVTGLSGG
ERRVALCKLLSQPLLSDLEPNHLDATIAWLEKHLGRDYPGAWMITHDYFLDN
VTGWIILEDRGPIPEGNYSAVLOKAKEMQOEAREDSRQKAI SREQWETASSPKA
ROTQSKARIRAYDELVEAENRRPQDAQIVIPVAERLGRVIVAEENLTKSYGDRVLE
NLTKLPGGGIVGIPNGAGKTLFRMITGQSQPDSGVTVGETVHLGVVDQSRDLE
AGDKTWEETSGNDLILKLFKEVNSPACGAFNFKGGDOOKVGNLSGGORNRHLA
KMKAGVNLLEDEPNLDLTETLALEAEAPGCAVITSHDRMFLDLRLATHLAF
EGDSHVEWFEQNFEDYKDKVRRLGPESIKPRGRVSYKRLTR"

```

```

Alignment Scores:
Pred. No.: 6 37e-19 Length: 10129
Score: 336.50 Matches: 180
Percent Similarity: 40.9% Conservatives: 107
Best Local Similarity: 25.7% Mismatches: 279
Query Match: 9.7% Indels: 140
DB: 1 Gaps: 26

```

US-10-620-914-45 (1-648) x AE008128 (1-10129)

```

Qy 5 ArgAspGlyArg-----ProAlaSerTyrThrLysLysAsnPheSerLeuGluLysLeu 22
Db 909 CGCGAAGCGCTACAAATCCGGCGCTATAGCGGGCAATTCGGCCTGATGCAAAATC 968
Qy 23 Lys----- 23
Db 969 CGCGATGCCAGCGCCGCTCCCTGGGCTTCGATGGCCCGCCCTCCGGCCTGTCGATGCC 1028
Qy 24 LeuSerSerMetLysAspAspLeuThrValLeuArgHisMetTrpPheGlySerLysLys 43
Db 1029 GAACCAATCTCAATACCGCGCTCAATATATTCGGCGCGCATGGTGTGGCGGGTAAC 1088
Qy 44 GlyAspAspHisAlaAlaArgLeuGluSer-----PheTyrGlyProGlnAlaAla 60
Db 1089 GACCGTGACAAATGCCGCTCCCTTATGCGCGGGTATTATTACGATGC-CAAGCGCAA 1147
Qy 61 AlaPheAlaAlaArgLeuAlaGluArgSerAsnLeuLeuTrpValAspLeuGlyGly 80
Db 1148 GGGCATGCTGACGCTTCGGGAA-----CTGACGGCCTCTGCGCTGT 1189
Qy 81 ThrGlyGluAsnValAspMetMetAlaAspTyrIleAspLeuAlaLysPheLysSerIle 100
Db 1190 GCTGGAGACCGCGCGAGCTGA--GCGGCTTTTGGAAATGTCGCGCTCCCTTCGTC 1247
Qy 101 TyrValValAspLeuCysHisSerLeuCysGluValAlaLysLysLysAlaLys 120
Db 1248 GTCAATCTGGGCTCGTCCCGAGGATCAACATCGCGTAAATAATCGATCGATTGCGAGA 1307
Qy 121 GlyTrpLysAsnValGlnValValGluAlaAspAlaCysGlnPheAlaProProGluGly 140
Db 1308 -----ACCTCGGGA 1316
Qy 141 ThrAlaThrLeuIleThrPheSerTyrSerLeuThrMetIlePro-PropheHisAsnVa 160
Db 1317 ATAGCCCCGGGATGACAGCTAGTGGTACGTGACACCTCGCCTTCAGCGGCGCGCC 1376
Qy 160 lIleAspGlnAlaCysSerTyrLeuSerGlnAspGlyLeuValGlyValAlaAspPheTy 180
Db 1377 GTTCTCTGGTGGCGGCTCGGCTGGTGGCGGCTGTTTCCCAATGCTCTTTCTTTT-- 1434
Qy 180 rValSerGlyLysTyrAspLeuProLeuArgGlnMetProTrpSerArgPhePheTr 200
Db 1435 -----CGCTCTCTGCGACCAACCCCTGTCATCAAACTGT----- 1470
Qy 200 pArgSerIlePheAspIleAspAsnIleAspIleGlyProGluArgAlaTyr----- 218
Db 1471 ----AGCGTGGCGCTGTAGAC-----GCCTCTCAATCGACTGAGAGGTGAACATATGACGA 1523
Qy 219 -----LeuGluGlnLysLeuGlu----ArgValTrpGluGlnAsnTh 231

```

```

Db 1524 GTCGGCAACCCAGACGCGCTTCAGCAAAACACGAAACGAAAGTCGCAATTCCTCCAGC 1583
Qy 231 rGlnGlySerIleProTyrValProTrpLeuArgAlaProTyrTyrValTrpIleGlyAr 251
Db 1584 ACAAGGCACTCTCCAAAGCGGCT----- 1608
Qy 251 gLeuProSerValGlyHisAlaLeuHisGluGluArgValGluArgProProMetPhePr 271
Db 1609 -----GTCCGAACGGTTTTTTCGGCGCTCTCTTTTCCG 1640
Qy 271 oPro-ThrPheLeuTyrThrGlnSerTrpGluAspProGluProAspMetGluValMetG 291
Db 1641 GCCTCGTC-----TATCCGCACTCTGGGAAGACCCCGAGATCGCATCGAAGCGATGG 1694
Qy 291 luIleAsnProLysAspThrValLeuThrLeuThrSerGlyGlyCysAsnAlaLeuAsnL 311
Db 1695 AGCTTCGGGAGGCCACCGCATCGTCACCATCGCTCCGGCTCCGGCGGCTGCNAACATGCTGCGCT 1754
Qy 311 euLeuValGlnGlyAlaGlyGlnValValSerValAspCysAsnProAlaGlnSerAlaL 331
Db 1755 ATCTCTCGGCAACCCCGCGCATCGATGTGTGGACCTCAACCCGACCATCATCGCGC 1814
Qy 331 euLeuGluLysLysValAlaIleGlnLeu----GluPheGluAspValTrpGlnL 350
Db 1815 TGAACAAGCTGAAGCTCGCTGCTCCGCCATCTGCCCGCCCATCAGGATGTGGTGGCC 1874
Qy 350 euPheGlyGluGlyValHisProArgIleGluLeu-TyrGluLysLysLeuAlaPro 369
Db 1875 ACTTCGGCGCGCGCGG-CACCCGCAACAGCGTGGTTATGACCGTTTCATCGCGCAG 1933
Qy 370 PheLeuSerGlnThrSerHisAsnPheTrpSerLys-----Arg 382
Db 1934 CATCTGGATGCCAGCAAGGCATACTGTGCAAGCGCACCTTTCGCGCGCGCTCGC 1993
Qy 383 LeuTrpTrpPheGlnHisGlyLeuTyrTyrGlnGlyGlyMetGlyLysLeuCysTrpVal 402
Db 1994 ATTTGGTGTTCACAGGAACATCTACCGGACCGGCTGCTCGCGCTTTTCATCGCGCC 2053
Qy 403 LeuGlnCysLeuAlaValValLeuGlyLysThrValLysArgLeuAlaAsnAla 422
Db 2054 GGCCATCATGTCGCGCTGCGCGGTGCAAACTACC-----GNAATGGCCAAAGC 2107
Qy 423 ProMetGluGluGlnArgArgLeuTrpAspSerAsnMetLeuIleHisPheValLys 442
Db 2108 CGGACGCTGACGACGACGCGCAGTTTGTGACAGCAAGTTCGCGCGCTTTTC----- 2161
Qy 443 AsnGlyProLysProLeuVal----TrpLeuPheValLysPheValSerLeuValLeuPhe 461
Db 2162 -----GACAAAGCGGTGGTGGCTGGCTGCGAAGCGCAAGAGCTCGCTT----- 2206
Qy 462 AsnLysAlaValLeuTrpPheGlyGlyValProGlyLysGlnTyr-----Ala 478
Db 2207 -----TTGCGCTTGGCATTTCCGCGCGCCCATGATGACGAGCTGGCA 2248
Qy 479 LeuIleLysAlaAspGlyIleProIleGluAsnTyrIleAlaArgThrMetAspGlyVal 498
Db 2249 AGCTTTTCCAGCGACGCGC---ACGGTTGCTCTCGCTCAAGGAGCGCTGGAAAAGCTT 2305
Qy 499 AlaGluAsnSerHisValArgLysGlnAsnTyrPheTyrTyrAsnCysLeuThrGlyLys 518
Db 2306 GCCTGCAACTTCCGCTC---AGCGCACAATATTTCGCTGGCAGGCGCTTTGCGCGCGCT 2362
Qy 519 PheLeuArgAspAsn-----CysProThrTyrLeuArgGluAlaAlaPheAlaThr 535
Db 2363 TATCCGAGCGCATGAGGTCGCTCCGCTTATCTCAAGCGCGCATATGACGAAAG 2422
Qy 536 LeuLysSerGlyValValAspAsnLeuThrValSerThrAsnPheMetGluGluLeu 555
Db 2423 ATCCGCAACAACCGCG---CGGTCGCGGTGTCATCAGCGCACCTATATACGAGCTGCTT 2479
Qy 556 Lys-----AlaArgThrTyrThrLysValIleLeuMetAspHisValAspTrpLeu 572
Db 2480 TCCGCAAGCGCGCAATATGCGTCGCTATATCTCTGTCGATGCGGAGCTGGATG 2539

```

Qy 573 AspMetProVallalaAenGluLeuAlaGluCysLeuAlaLysGlnValAlaProGlyGly 592
 Db 2540 ACCGATGTCAGCTCAACGAGTTATGTCGCAGATCAGCGCATCGCGGCGCA 2599
 Qy 593 IleValIleTrpArgSerAlaSerLeu-----SerProProTyrAlaGlu 607
 Db 2600 CGGTCATCTCCGACCGCGCGCGA-AAAGAGCGTTATCGAGGCGCGGTTTCGCCCGA 2658
 Qy 608 LeuIleGlnLysAlaGlyPheAspValArgCysIleArgAlaThrGlnGlyTyrMet 627
 Db 2659 CATCCGCAACAGTGGGTCTATCTCGAAGAGCGCTCCACGA---ACTCAAC--GCCATG 2713
 Qy 628 AspArgValAsnMetTyrSerSerPheTyrMetAlaArgLysGlyAlaLysLys 646
 Db 2714 GACCGCTCGGCCATTTATGCGGCTTCATATCACCAGAGGCTATGGCATGAAA 2770

RESULT 14
 LOCUS AE009162 10134 bp DNA linear BCT 28-MAY-2004
 DEFINITION Agrobacterium tumefaciens str. C58 circular chromosome, section 188
 of 256 of the complete sequence.
 ACCESSION AE009162 AE008688
 VERSION AE009162.1 GI:17740578
 KEYWORDS Agrobacterium tumefaciens str. C58
 SOURCE Agrobacterium tumefaciens str. C58
 ORGANISM Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 Rhizobiaceae; Agrobacterium.
 REFERENCE 1 (bases 1 to 10134)
 Wood, D.W., Setubal, J.C., Kaul, R., Monks, D., Chen, L., Wood, G.E.,
 Chen, Y., Woo, L., Kitajima, J.P., Okura, V.K., Almeida, N.F., Jr.,
 Zhou, Y., Bovee, D., Sr., Chapman, P., Clendenning, J., Deatherage, G.,
 Gillet, W., Grant, C., Guenther, D., Kutyavin, T., Levy, R., Li, M.,
 McClelland, E., Palmieri, A., Raymond, C., Rouse, G., Paulsen, I.,
 Saenphimmachak, C., Wu, Z., Gordon, D., Eisen, J.A., Paulsen, I.,
 Karp, P., Romero, P., Zhang, S., Yoo, H., Tao, Y., Biddle, P., Jung, M.,
 Krespan, W., Perry, M., Gordon-Kamm, B., Liao, L., Kim, S., Hendrick, C.,
 Zhao, Z., Dolan, M., Tingey, S.V., Tomb, J., Gordon, M.P., Olson, M.V.
 and Nester, E.W.
 The genome of the natural genetic engineer Agrobacterium
 tumefaciens C58
 Science 294 (5550), 2317-2323 (2001)
 11743193
 2 (bases 1 to 10134)
 Wood, D.W., Setubal, J.C., Kaul, R., Monks, D., Chen, L., Wood, G.E.,
 Chen, Y., Woo, L., Kitajima, J.P., Okura, V.K., Almeida, N.F., Jr.,
 Zhou, Y., Bovee, D., Sr., Chapman, P., Clendenning, J., Deatherage, G.,
 Gillet, W., Grant, C., Guenther, D., Kutyavin, T., Levy, R., Li, M.,
 McClelland, E., Palmieri, A., Raymond, C., Rouse, G.,
 Saenphimmachak, C., Wu, Z., Gordon, D., Eisen, J.A., Paulsen, I.,
 Karp, P., Romero, P., Zhang, S., Yoo, H., Tao, Y., Biddle, P., Jung, M.,
 Krespan, W., Perry, M., Gordon-Kamm, B., Liao, L., Kim, S., Hendrick, C.,
 Zhao, Z., Dolan, M., Tingey, S.V., Tomb, J., Gordon, M.P., Olson, M.V.
 and Nester, E.W.
 Direct Submission
 Submitted (27-SEP-2001) Department of Microbiology, University of
 Washington, 1959 NE Pacific Ave, Box 357242, Seattle, WA
 98195-7242, USA
 TITLE Location/Qualifiers
 JOURNAL 1. 10134
 PUBMED /organism="Agrobacterium tumefaciens str. C58"
 REFERENCE /mol_type="genomic DNA"
 AUTHORS /strain="C58"
 /isolate="U. Washington"
 /db_xref="taxon:176299"
 99. 1175
 /genes="Atu2117"
 99. 1175
 /gene="Atu2117"
 /note="identified by sequence similarity; putative; ORF
 located using Blastx/Glimmer"
 /codon_start=1

gene
 CDS
 /transl_table=11
 /product="lytic murein transglycosylase"
 /protein_id="AAL43107.1"
 /db_xref="GI:17740579"
 /translation="MNSGKRGSRALLSSSTVGLALALASCTSVAAKEGPNPKI
 AVTSSQAQTSQPAATAEAVAGATPAAPAVOTADAAASQAAALPTAVAVKGRVAL
 PPAEIAAANAATQOQTEVAGAAATTAATTAALVETHHTTPAAALTAETANAAK
 PAMNAADVLQVAAVKGSPFAAPPSPGSGITELTAERKVIPLPKPDSPTVLAAG
 PTNAALAIQNESMTAPARTAPAGREKLSGLITKYATMYDVEDLVHRYVQRESRYN
 PGAYSGNFGMQIRHATARSLSGLGDPASGLFDAETNLKYAVKYLARGAVVAGNDRDN
 AVLYARGVYDAKRKGMHLVLRN"
 complement (1274..1516)
 /genes="Atu2118"
 complement (1274..1516)
 /genes="Atu2118"
 /note="putative; ORF located using Glimmer"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAL43108.1"
 /db_xref="GI:17740580"
 /translation="MFTSOSIERLQPPRYSLTMGVAGRESEKKRPLENQAPAPARAGH
 QRTAAPLEGVVTDSHVITPGAIPVQLIDFLRHV"
 1517. 2767
 /genes="Atu2119"
 1517. 2767
 /genes="Atu2119"
 /note="identified by sequence similarity; putative; ORF
 located using Blastx/Glimmer"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAL43109.1"
 /db_xref="GI:17740581"
 /translation="MTPSAKPTGFSKNTKLKALLQKALSLSGLSEFFGVLFGSLV
 YQIWEDEPIDMEAMELGHRIVTIGSGCNMLAYLSRNPASIDVVDLNPHEIALNK
 LKLAARHPHQAQDVVRFGRAGTSNSGYDFIAEHLDAITTKAYSKRSLGRRRI
 VFNDRITYTGLLGFIGHIMARLHVKLTEMAKRTLDQEQFDSKVAFDFRP
 SVWRITKRSLSLGLGIPPRQYDELASLSGDSGTVASVLEKLEKACNPFPSNDYFAP
 QAFAPRYPEHEGALPAYLKPEYVEKIRNNTARVAVHATYTELLSRKPANGVDRIYL
 LDAQDMTDVQLNEGLWSQISRTAASGARVIFRTAAEKSVIEGRSLSDIRNQWYLEER
 SNELNMDRSALYGGFHIYQRAMA"
 2764. 3483
 /genes="Atu2120"
 2764. 3483
 /genes="Atu2120"
 /note="identified by sequence similarity; putative; ORF
 located using Blastx/Glimmer"
 /codon_start=1
 /transl_table=11
 /product="O-methyltransferase"
 /protein_id="AAL43110.1"
 /db_xref="GI:17740582"
 /translation="MKTIGENVGLADSAHAGLMDRMYRQRIHYDITRKYLLGRDRTP
 ISGLDVPGKGTLLIGCGTGRNLLASRRFPDPAKFGLDISAEMLLTASENFAGKRI
 PILRVADATAPRSSESFGPDGFVIMIPYALSMIPDWEKAIQEAALAKPGSGSLHVD
 FQOQELPKWFTLLQWLTRFHVTFRANLRYVLANAGRFDCNLVFEETARGVAMEA
 VITLPVAEAPQPKIHRLLADA"
 3487. 4498
 /genes="lyc"
 /note="synonym: Atu2121"
 3647. 4498
 /genes="lyc"
 /note="identified by sequence similarity; putative; ORF
 located using Blastx/Glimmer"
 /codon_start=1
 /transl_table=11
 /product="glycosyl hydrolase/lysozyme"
 /protein_id="AAL43111.1"
 /db_xref="GI:17740583"
 /translation="MRKSGDTRARRFAAGAWATIIIRVPMRLRFSALLSALLVSGGCTS
 TSYDLLETASVSRKPFSDYDQDFGVNPHRHEHGDVSKWNGDWDQVTRKSGVSF
 VFIKATEGSDRIDPKFGDHWRSAAANILHAPVHYFYFCSTADQADWFISNPKRAV"

1641	Db	GCCTCGTC-----TATCCGCAGATCTGGGAAGACCCCGAGATCGACATCGAAGCAGCATGG	1694
291	Qy	luLeuAsnProLysAspThrValLeuThrLeuThrSerGlyGlyCyAsnAlaLeuAsnL	311
1695	Db	AGCTTGGCGAAGGCCACCGCATCGTACCATCGCTCCGGCGGCTGCAACATGCTGGCCT	1754
311	Qy	euLeuValGlnGlyAlaGlyGlnValSerValAspCysAsnProAlaGlnSerAlaL	331
1755	Db	ATCTCTCGCCCAACCCCGCCAGCATGATGTGGTGAGCTTCAACCCCGCACCATCGCGC	1814
331	Qy	euLeuGluLeuLysValAlaIleGlnGlnLeu--GluPheGluAspValTrrpGlnL	350
1815	Db	TGAACAGCTGAAGCTCGCTCGCTTCGCCCATCTGCCCGCCCATCAGAGTGTGGTGC	1874
350	Qy	eupheGlyGluGlyValHisProArgIleGluGluLeu-TyrGluLysLysLeuAlaPro	369
1875	Db	ACTTCGCCCGCGCGG-CACCCGCGACCAACGCTCGGTTATGACCGTTTTCATCGCGAG	1933
370	Qy	PheLeuSerGlnThrSerHisAsnPheTrrpSerLys-----Arg	382
1934	Db	CATCTGATGTCACGACCAAGGCATACTGTCGAAGCGCACCCCTTTCGCCGCCCGCTCGC	1993
383	Qy	LeuTrrpTyrPheGlnHisGlyLeuTyrTyrGlnGlyMetGlyLysLeuCysTrrpVal	402
1994	Db	ATTTCGGTGTTCGACAGAACATCTACCGGACCGGCTGCTCGGCGCTTTCATCGCGGCC	2053
403	Qy	LeuGlnCysLeuAlaValLeuGlyLeuGlyLysThrValLysArgLeuAlaAsnAla	422
2054	Db	GGCCACATATGCGCCGCTGCACGGGTGAACTCACC-----GAAATGGCCAAAGACC	2107
423	Qy	ProThrMetGluGlnArgLeuTrrpAspSerAsnMetLeuIleHisPheValLys	442
2108	Db	CGGACGCTGGACCAACAGCCGCAAGTTTTTGACAGCAAGGTCGCGCGCTTTTC-----	2161
443	Qy	AsnGlyProLysProLeuVal---TrrpLeuPheValLysPheValSerLeuValLeuPhe	461
2162	Db	-----GACAAAGCCGGTGGTGGCTGGCTGACGAAGCGCAAGAGTCGCTT-----	2206
462	Qy	AsnLysAlaValLeuTrrpPheGlyGlyValProGlyLysGlnTyr-----Ala	478
2207	Db	-----TTCCGGCTTGGCATTTCCGCGCGCCAGTAGTACGAGCTGGCA	2248
479	Qy	LeuIleLysAlaAspGlyIleProIleGluAsnTyrIleAlaArgThrMetAspGlyLys	498
2249	Db	AGCCTTCCAGCGACGCGC---ACGGTTCCTCCGCTCTCAAGGAGCGGCTGGAAAAAGCTT	2305
499	Qy	AlaGluAsnSerHisValArgLysGlnAsnTyrPheTyrTrrpAsnCysLeuThrGlyLys	518
2306	Db	GCTTGCACAACTTCGCGCTC---AGCGCAAAATATTTTCGCTGGCAGGCGCTTTCGCGCC	2362
519	Qy	PheLeuArgAspAsn-----CysProThrTyrLeuArgGluAlaAlaPheAlaThr	535
2363	Db	TATCCCGCGCGCATGAGGTGCTCGCTTATCTCAAGCGGAATATTACGAAAAG	2422
536	Qy	LeuLysSerGlyValValAspAsnLeuThrValSerThrAsnPhePheMetGluGluLeu	555
2423	Db	ATCCGCAACAAACCCGCG---CGCGTCGCGGTGCATCACGCCACCTATATCCGAGCTGCTT	2479
556	Qy	Lys-----AlaArgThrTyrThrLysValIleLeuMetAspHisValAspTrrpLeu	572
2480	Db	TCCCGCAAGCCGCAATATGCGTCGACCGCTATATCTTCCTGCTCGCAGGACTCGATG	2539
573	Qy	AspMetProValAlaAsnGluLeuAlaGluCysLeuAlaLysGlnValAlaProGlyGly	592
2540	Db	ACGGATGTGCAGTCAACAGATTATGTCGCAGATCAGCGCGCACTCCGCATCCGGGGCA	2599
593	Qy	IleValIleTrrpArgSerAlaSerLeu-----SerProTrrpTrrpAlaGlu	607
2600	Db	CGCGTCATCTTCGCACCGCGGCGCA-AAAGAGCGTTATCGAGGCGCGCTTTCGCCCA	2658
608	Qy	LeuIleGlnLysAlaGlyPheAspValArgCysIleAArgAlaThrGlnGlyTyrMet	627
2659	Db	CATCCGCAACCACTGGTGTCTATCTCGAAGAGCCCTCCAAACGA---ACTCAAC---GCCATG	2713

BA000012_63 6300001 6410000
BA000012_64 6400001 6510000
BA000012_65 6500001 6610000
BA000012_66 6600001 6710000
BA000012_67 6700001 6810000
BA000012_68 6800001 6910000
BA000012_69 6900001 7010000
BA000012_70 7000001 7036071
Continuation (14 of 71) of BA000012 from base 1300001 (BA000012 Mesorhizobium loti MAF3)

Alignment Scores:
Pred. No.: 6, 31e-17 Length: 110000
Score: 329.00 Matches: 118
Percent Similarity: 42.8% Conservative: 66
Best Local Similarity: 27.4% Mismatches: 184
Query Match: 9.5% Indels: 63
DB: 1 Gaps: 15

US-10-620-914-45 (1-648) x BA000012_13 (1-110000)

Qy 248 TrpIleGlyArgLeuProSer-----ValGlyHisAlaLeuHisGluGluArgValGlu 265
Db 13624 TGGAAAGCGCGTCTACCAAGCGCGCTTTCCAAAGCGGCATCTCCGAGCGGTGT 13683
Qy 266 ArgProMetPheProPro-ThrPheLeuTyrThrGlnSerTrpGluAspProGluPr 285
Db 13684 CGCCTTCCTGTTTCCGGCTCGTC-----TATCCGACATCTGGGAAGACCCCGATGT 13737
Qy 285 oAspMetGluValMetGluLeuAsnProLysAspThrValLeuThrLeuThrSerGlyG1 305
Db 13738 CGACATGGAGCGCATCGAGCTTGGTCAGGGCCATCGCATCGTCACATCGCTTCGCGCG 13797
Qy 305 YCysAsnAlaLeuAsnLeuValGlnGlyAlaGlyGlnValValSerValAspCysAs 325
Db 13798 CTGCAACATCTCGCTTACCTACCCCTTCGCGCGCACGATCGACCGCTCGACCTCAA 13857
Qy 325 nProAlaGlnSerAlaLeuLeuGluLeuLysValAlaLeuLeuGlnGluPheG1 345
Db 13858 CGCGGCCCAATCGCGCTGAACCGCATGAAGCTGGAGCGGTGGCGGTCTGCGCTCGCA 13917
Qy 345 u---AspValTrpGlnLeuPheGlyGluGlyValHisProArgIleGluLeuTyrG1 364
Db 13918 GGGCGATCTGTCGCTTTTTCGGCGCGCGCGCACCGACCAATTCGCAAGCTATGA 13977
Qy 364 uLysLysLeuAlaProPheLeuSerGlnThrSerHisAsnPheTrpSerLysArgLeuTr 384
Db 13978 CCGCTTTATTTCGGCGCATCTCGATCGCGTCAGCGCCACTATTGGGAGCGCGCAACTG 14037
Qy 384 p-----TyrPheGlnHisGlyLeuTyrTyrGlnGlyGlyMetG1 397
Db 14038 CGGTGTCGGCGCGCATCGCGTCTTCGACCGCAATTTCTACAGACCGCGCTGCTCGG 14097
Qy 397 YLysLeuCysTrpValLeuGlnCysLeuAlaValValLeuGly----- 411
Db 14098 C-----CTGTTTCATCGCCATGGCCATCGCCATCGCAGCGGAAAT 14133
Qy 412 -LeuGlyLysThrValLysArgLeuAlaAsnAlaProThrMetGluGlnArgArgLe 431
Db 14134 CTTTCGGCGTCAACCGCGCCCATGATGGAAGCGGAATATCGCGCAGCAGCGCGCTT 14193
Qy 431 uTrpAspSerAsnMetLeuIleHisPheValLysAsnGlyProLysProLeuValTrpLe 451
Db 14194 CTTTCACAGAGAGTGGCGCGGTCTTCGACAG-----AAGCTTTGAATGGGC 14244
Qy 451 uPheValLysPheValSerLeuValLeuPheAsnLysAlaValLeuTrpPheGlyG1 471
Db 14245 GACCTCGGTAAGCGCTCGCTG-----TTTCGGCTCGG 14277
Qy 471 YValProGlyLysGlnTyr-----AlaLeuIleLysAlaAspGlyIleProIleGluAs 489
Db 14278 CATTTCGCGCGCGAGTACGATTCCTCATCCCTCAGGCGCGCATCGCCATGGCCAGCGT 14337
Qy 489 nTyrIleAlaArg-----ThrMetAspGlyValAlaGluAsnSerHisValAr 505

Search completed: March 14, 2006, 05:00:33
Job time : 9713 secs

Db 14338 TCTGAAGCGCGCTCGCTGCGAATTTTCCCTGGAATC----- 14386
Qy 505 gLysGlnAsnTyrPheTyrTyrAsnCysLeuThrGlyLysPheLeuArgAspAsn----- 523
Db 14387 -----AATTATTTCGCTCGCGAGGCTTTTTCGCCCGCGCTATCCAAATCCCGGTAGGC 14439
Qy 524 -----CysProThrTyrLeuArgGluAlaAlaPheAlaThrLeuLysSerGlyValValAs 542
Db 14440 CGCCCTGCGCGCTATCTGGAAGAGAGAACTACGAACCATCCGC-----GGCAATATCGA 14496
Qy 542 pAsnLeuThrValSerThrAsnPhePheMetGluGluLeu-----LysAlaArgTh 559
Db 14497 CGCGTTCGCGCATCCACCATGCCAATCTGATCGAATTTCTCGCGCGCAAGACGCGGCAC 14556
Qy 559 rTyrThrLysValIleLeuMetAspHisValAspTrpLeuAspMetProValAlaAsnG1 579
Db 14557 CGTCGATCGCTTCATCTGCTCGATGCGCAGGAGTGGATGACCATGACCATGACGCTCAACGC 14616
Qy 579 uLeuAlaGluCysLeuAlaLysGlnValAlaProGlyGlyIleValIleTrpArgSerAl 599
Db 14617 GCTGTGTCGGAATCAGCCGACCGCTTCGCGAGCGCGCGCTCATCTTCGCGCACCGC 14676
Qy 599 a-----SerLeuSerProProTyrAlaGluLeuIleGlnLysAlaGlyPheAsp-V 616
Db 14677 CGCGGAGCGCGCTGCTGCGCAGCGCGCTCTCGACCTGCTGCTCGACCATGGGACTA 14736
Qy 616 alArgCysIleArgArgAlaThrGlnGlyTyrMetAspArgValAsnMetTyrSerSerP 636
Db 14737 TCAGGACGAGCGCTCGCGGAAATTCGCGCACGC-GACCGTTTCGGCCATCTATGCGCGCT 14795
Qy 636 heTyrMetAlaArgArgLysGlyAla 644
Db 14796 TCACCTCTATGTGAAGCGCACGCA 14821

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 13, 2006, 21:06:21 ; Search time 9929 Seconds
(without alignments)
11146.561 Million cell updates/sec

Title: US-10-620-914-44
Perfect score: 1947
Sequence: 1 atggggctgggtctgcagg.....gogccaagaaggacaactaa 1947

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*
1: gb_ba.*
2: gb_in.*
3: gb_env.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pr.*
9: gb_ro.*
10: gb_sts.*
11: gb_sy.*
12: gb_un.*
13: gb_vi.*
14: gb_htg.*
15: gb_pl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1883	96.7	2001	15	AY656806 Chlamydom
2	100.6	5.2	110000	15	Continuation (13 o
3	100.2	5.1	3045	1	AF329857 Rhodobact
4	87.4	4.5	110000	1	Continuation (17 o
5	83.8	4.3	110000	15	Continuation (21 o
6	82.2	4.2	125020	8	AF429315 Homo sapi
7	82	4.2	31065	15	AC152119 Pichia st
8	80	4.1	110000	1	Continuation (64 o
9	77.2	4.0	110000	15	Continuation (21 o
10	76.8	3.9	349260	1	BX572595 Rhodosphe
11	76	3.9	1932	1	AJ605554 Nonomurae
12	75.6	3.9	137241	13	AY386263 Orf virus
13	74	3.8	110000	1	Continuation (20 o
14	73	3.7	75337	15	BX842626 Neurospor
15	72.6	3.7	110000	1	Continuation (25 o
16	72.4	3.7	137560	6	CQ830057 Sequence
17	72.4	3.7	137560	6	AX754989 Sequence
18	70.8	3.6	110000	1	Continuation (14 o

c	19	70.4	3.6	314100	1	SC0939106	AL939106 Streptomy
	20	70	3.6	294800	1	SME591789	AL591789 Sinorhizo
	21	69.8	3.6	305584	1	AE016920	AE016920 Chromobac
c	22	69.6	3.6	300100	1	SC0939123	AL939123 Streptomy
	23	69.4	3.6	12370	1	AE004680	AE004680 Pseudomon
c	24	69.4	3.6	189050	1	AL646086	AL646086 Ralstonia
	25	69.2	3.6	1500	13	AF294740	AF294740 Baboon he
	26	69	3.5	2563	1	AY228174S1	AY228174 Streptomy
	27	68.6	3.5	110000	14	CT005248_1	Continuation (2 of
	28	68.4	3.5	66669	1	AME16952	Y16952 Amycolatops
	29	68.2	3.5	2440	1	SLGROESEL	X95970 S. lividans
	30	68.2	3.5	2830	2	AF027972	AF027972 Nephila c
	31	68.2	3.5	2830	6	AR091362	AR091362 Sequence
	32	67.8	3.5	1092	1	AB046667	AB046667 Streptomy
	33	67.8	3.5	1092	6	BD095293	BD095293 Genes of
	34	67.8	3.5	1092	6	BD141256	BD141256 Isopenten
	35	67.8	3.5	1092	6	E48856	E48856 Gene of enz
	36	67.8	3.5	1104	6	BD097684	BD097684 A method
	37	67.8	3.5	1104	6	BD141277	BD141277 Isopenten
	38	67.8	3.5	6798	1	AB037666	AB037666 Streptomy
	39	67.8	3.5	6798	6	BD095289	BD095289 Genes of
	40	67.8	3.5	6798	6	BD097685	BD097685 A method
	41	67.8	3.5	6798	6	E48852	E48852 Gene of enz
	42	67.8	3.5	30837	1	AJ843080	AJ843080 Streptomy
	43	67.8	3.5	39237	1	AB211959	AB211959 Streptomy
c	44	67.8	3.5	42149	1	AJ786317	AJ786317 Streptomy
	45	67.8	3.5	139962	13	AY386264	AY386264 Orf virus

ALIGNMENTS

RESULT 1
AY656806
LOCUS Chlamydomonas reinhardtii betaine lipid synthase (BTA1) mRNA,
DEFINITION complete cds.
ACCESSION AY656806
VERSION AY656806.1 GI:50261571
KEYWORDS Chlamydomonas reinhardtii
SOURCE Chlamydomonas reinhardtii
ORGANISM Chlamydomonas reinhardtii
REFERENCE 1 (bases 1 to 2001)
AUTHORS Riekhof, W.R., Sears, B.B. and Benning, C.
TITLE Annotation of Genes Involved in Glycerolipid Biosynthesis in Chlamydomonas reinhardtii: Discovery of the Betaine Lipid Synthase BTA1C
JOURNAL Eukaryotic Cell 4 (2), 242-252 (2005)
PUBMED 15701786
REFERENCE 2 (bases 1 to 2001)
AUTHORS Riekhof, W.R. and Benning, C.
TITLE Direct Submission
JOURNAL Submitted (17-JUN-2004) BMB, MSU, Biochemistry 215, East Lansing, MI 48824-1319, USA

FEATURES
source
1..2001
/organism="Chlamydomonas reinhardtii"
/mol_type="mRNA"
/strain="CC125"
/db_xref="taxon:3055"
1..2001
/gene="BTA1"
1..2001
/gene="BTA1"
/note="multifunctional protein with S-adenosylmethionine:diacylglycerol 3-amino-3-carboxypropyl transferase and S-adenosylmethionine:diacylglycerolhomoserine N-methyltransferase activities"
/codon_start=1
/product="betaine lipid synthase"

CDS
/gene="btaB"
1791..2423
/gene="btaB"
/codon_start=1
/transl_table=11
/product="S-adenosylmethionine:diacylglycerolhomoserine-N-methyltransferase"
/protein_id="AAK53561.1"
/db_xref="GI:14091031"
/translation="MTDTHAALMDATYRQRIYDVTRRHFLGLRDLIAELDPPPG
ARVLEIACGTGRNLDIGRRPGCSGLDISQEMLASARARIGREATLALGDATFE
ALPLFGDTGFERIVLGYALSMI PMREALREAHLPVGRHLVDFGQAGLPGAR
AGURGWIGRIVTPRDDLTALGETALGIGVAYEYRSLGGVAILTLTR"
2687..2762
/product="tRNA-Pro"

tRNA
ORIGIN

Query Match 5.1%; Score 100.2; DB 1; Length 3045;
Best Local Similarity 50.9%; Pred. No. 0.0092;
Matches 299; Conservative 0; Mismatches 273; Indels 15; Gaps 2;
QY 809 TCCCGCCACCTTCCTGTACACGAGTCGTGGGAGGACCCGAGCGGATATGAGGTGA 868
DB 665 TCTTCCACGGCCTCGTCTATCCGAGATCTGGGAGGATCCGGCGGTGGACATGGCGGCC 724
QY 869 TGGAGATCAACCCAGGACACGCTGCTGACCCCTGACTAGCGGGGTGCAATGCCCTGA 928
DB 725 TCGCCATCCCGCCCGGGGACCGGCTGTGTGCCATCGCTCGGGCGGTGTGCAACGTGCTTT 784
QY 929 ACCTGTGTGTGACGGGGCGCGCAGGTGTGTGCTGGTGAAGTCAACCCCGCGCAGTCGG 988
DB 785 CCTATCTCAGCAGGGCGCGGCTCGATCTTCGCGGTGGATCTTCGCGCGCCCATGTGG 844
QY 989 CGTTCTGAGCTGAAGAGGTGGCCATTCAGCAGCTG---GAGTTTGAAGACGTGTGGC 1045
DB 845 CGCTGGGGCGGCTGAAGCTCGCGCGCGCGGACGCTGCCCGACCATGCCCGCTTCTTCG 904
QY 1046 AGCTGTTCGGGAGGGCGTGCACCCGCGCATTCGAGAGCTGTACGAGAGAGTGGCGC 1105
DB 905 ATCTTTCGTCGCGCAGACCTGCGCGGCAATGCGGCGCTCTACGACCGCCCATCGCGC 964
QY 1106 CCTTCTGTGTCGAACACAGCACAACCTTCTGTGTCGAAGCGCTCTGTACTTCCAGCACG 1165
DB 965 CCGGCTTCGACGGCGCGGCGCGCTACTTGGGAGGCGCGAGCCCTTCGGCGGGGCA 1024
QY 1166 GCCTGTACTACGAGGGCGGATGGGCAAGCTGTGTGGTGTGTGCTGAGTGTCTGGCGG--- 1222
DB 1025 TCCAGCTGTTCGAGCGCGGCTTCTACCGGACGCTGCTCGCGCGCTTCATCGGCGCG 1084
QY 1223 -----TGCTGTGGACTGGCAAGACCGTCAAGCGCTTCGCCAAGCGCGCCACAA 1273
DB 1085 CCATAGCTTCGGCGGGCGCGGGGACCGACCTTCGGGGCTTCTGCACTGTCCGACA 1144
QY 1274 TGGAGGACGACGCGCGCTCTGTGGGACAGCAACATGCTTCATCCACTTCGTGAAGAACGGC 1333
DB 1145 TCGAGGCGCAGCGCAGCTTCTTACGCCCATATCGGGCGGCTCTTCGAGGCGCGCGTGG 1204
QY 1334 CCAAGCGCTGTGTGTGCTGTGTGTAAGTTCTGTGACCTGTGTGCTC 1380
DB 1205 TGCAGGCGCTTCGCCGACGCGCGCGCGCTCTTCGGGCTGGGGATC 1251

RESULT 4
AP006618_16
WPCOMMENT
Sequence split into 61 fragments LOCUS AP006618 Accession AP006618
Fragment Name Begin End
AP006618_00 1 110000
AP006618_01 100001 210000
AP006618_02 200001 310000
AP006618_03 300001 410000
AP006618_04 400001 510000
AP006618_05 500001 610000

AP006618_06 600001 710000
AP006618_07 700001 810000
AP006618_08 800001 910000
AP006618_09 900001 1010000
AP006618_10 1000001 1110000
AP006618_11 1100001 1210000
AP006618_12 1200001 1310000
AP006618_13 1300001 1410000
AP006618_14 1400001 1510000
AP006618_15 1500001 1610000
AP006618_16 1600001 1710000
AP006618_17 1700001 1810000
AP006618_18 1800001 1910000
AP006618_19 1900001 2010000
AP006618_20 2000001 2110000
AP006618_21 2100001 2210000
AP006618_22 2200001 2310000
AP006618_23 2300001 2410000
AP006618_24 2400001 2510000
AP006618_25 2500001 2610000
AP006618_26 2600001 2710000
AP006618_27 2700001 2810000
AP006618_28 2800001 2910000
AP006618_29 2900001 3010000
AP006618_30 3000001 3110000
AP006618_31 3100001 3210000
AP006618_32 3200001 3310000
AP006618_33 3300001 3410000
AP006618_34 3400001 3510000
AP006618_35 3500001 3610000
AP006618_36 3600001 3710000
AP006618_37 3700001 3810000
AP006618_38 3800001 3910000
AP006618_39 3900001 4010000
AP006618_40 4000001 4110000
AP006618_41 4100001 4210000
AP006618_42 4200001 4310000
AP006618_43 4300001 4410000
AP006618_44 4400001 4510000
AP006618_45 4500001 4610000
AP006618_46 4600001 4710000
AP006618_47 4700001 4810000
AP006618_48 4800001 4910000
AP006618_49 4900001 5010000
AP006618_50 5000001 5110000
AP006618_51 5100001 5210000
AP006618_52 5200001 5310000
AP006618_53 5300001 5410000
AP006618_54 5400001 5510000
AP006618_55 5500001 5610000
AP006618_56 5600001 5710000
AP006618_57 5700001 5810000
AP006618_58 5800001 5910000
AP006618_59 5900001 6010000
AP006618_60 6000001 6021225
Continuation (17 of 61) of AP006618 from base 1600001 (AP006618 Nocardia farcinica IFM 10

Query Match 4.5%; Score 87.4; DB 1; Length 110000;
Best Local Similarity 47.2%; Pred. No. 0.12;
Matches 342; Conservative 0; Mismatches 371; Indels 12; Gaps 2;
QY 800 CGCCCATGTTCCCGCCACCTTCTGTACACGACGCTGTGGGAGGACCCCGAGCCGGATA 859
DB 27308 CGCGGTGACCATGCTGCTGTCTACAGACCTTCGACGAGGATTCGCTCGGAGT 27367
QY 860 TGGAGTGTATGGAGATCAACCCCAAGACACGCTGTGTGACCTGTACTAGCGGCGGTGCA 919
DB 27368 TGC GCGCTGACCATCGGCGAGCGACGACGCTGTGGCGGTGACCGAAGCGGTTGCC 27427
QY 920 ATGCCCTGAACCTGCTGGTTCAGGGGGCCGCGGAGGTGTGCTGGTGTGACTGCAACCCCG 979
DB 27428 GCGCACTGAGTCTGGTGGTGCAATCCGCGCTCGGTGCTGCTCGGTGAGAGCTCCGCCG 27487

QY 395 CTGCGCAATTTGGCCCTGAGGGGACCGGAGGCTCATCACCTTCTCTACTCGCTCA 454
DB 20341 CCTGTGACTTTGAATCGACTACAGACTGCTGACTTTGATCAGCTTTTCTACTCTCTTT 20400
QY 455 CGATGATTCCACGGTTCACCAAGTCAATCGACGAGGCTTCTGCTACTCTTCCAGACG 514
DB 20401 CTATGATTCCACTTTCAATGCTGCTATCGAATGATGATCACAAGTTGTAAGCAGG 20460
QY 515 GCCTGTGGGGCTTGGCGACTT 536
DB 20461 GTATTATTGCTGTGTAGACTT 20482

RESULT 8
BA000040_63/c
WPCOMMENT

Sequence split into 91 fragments LOCUS BA000040 Accession BA000040

Fragment Name	Begin	End
BA000040_00	1	110000
BA000040_01	100001	210000
BA000040_02	200001	310000
BA000040_03	300001	410000
BA000040_04	400001	510000
BA000040_05	500001	610000
BA000040_06	600001	710000
BA000040_07	700001	810000
BA000040_08	800001	910000
BA000040_09	900001	1010000
BA000040_10	1000001	1110000
BA000040_11	1100001	1210000
BA000040_12	1200001	1310000
BA000040_13	1300001	1410000
BA000040_14	1400001	1510000
BA000040_15	1500001	1610000
BA000040_16	1600001	1710000
BA000040_17	1700001	1810000
BA000040_18	1800001	1910000
BA000040_19	1900001	2010000
BA000040_20	2000001	2110000
BA000040_21	2100001	2210000
BA000040_22	2200001	2310000
BA000040_23	2300001	2410000
BA000040_24	2400001	2510000
BA000040_25	2500001	2610000
BA000040_26	2600001	2710000
BA000040_27	2700001	2810000
BA000040_28	2800001	2910000
BA000040_29	2900001	3010000
BA000040_30	3000001	3110000
BA000040_31	3100001	3210000
BA000040_32	3200001	3310000
BA000040_33	3300001	3410000
BA000040_34	3400001	3510000
BA000040_35	3500001	3610000
BA000040_36	3600001	3710000
BA000040_37	3700001	3810000
BA000040_38	3800001	3910000
BA000040_39	3900001	4010000
BA000040_40	4000001	4110000
BA000040_41	4100001	4210000
BA000040_42	4200001	4310000
BA000040_43	4300001	4410000
BA000040_44	4400001	4510000
BA000040_45	4500001	4610000
BA000040_46	4600001	4710000
BA000040_47	4700001	4810000
BA000040_48	4800001	4910000
BA000040_49	4900001	5010000
BA000040_50	5000001	5110000
BA000040_51	5100001	5210000
BA000040_52	5200001	5310000
BA000040_53	5300001	5410000
BA000040_54	5400001	5510000

Continuation (64 of 91) of BA000040 based on base 6300001 (BA000040 Bradyrhizobium japonicum)

Query Match 4.1%; Score 80; DB 1; Length 110000;
Best Local Similarity 45.6%; Pred. No. 0.86;
Matches 379; Conservative 0; Mismatches 435; Indels 18; Gaps 2;

QY 972 CAACCCCGCGAGTCCGGCTTCTGAGCTGAAGAAGTGGCCATTTGAGCAGCTGGAGTT 1031
DB 73879 CATGCTCGAGACCACCGAGCGCGCTGGAGAACCAAGATCAACAGTCCGAGAAGAT 73820
QY 1032 TGAGGACGTGTGCGAGCTGTTGGCGAGGGGCTGACCCGCGCATTTGAGGAGCTGTACGA 1091
DB 73819 GGAGACGGTCGGCCAGCTCGCGCGCGCATTCGCCACAGCTTCAACAACGTGTGTCCGC 73760
QY 1092 GAAGAAGCTGGGCGCCCTTCTGTGCAAAACCAAGCCACAACTTCTGTGTCGAAGCGCTCTG 1151
DB 73759 CATCATGATGGCGAAGACTTCTGTGTAACGGCGCACAGCCGACCGATCCGTCTCCA 73700
QY 1152 GTACTTCCAGCAGCGCTGTACTACAGGGGCGCATGGGCAAGCTGTGTGGTGTGCA 1211
DB 73699 CGACATCATGCAGATCAAGCAGAACCGACGCGCGCGCATGCTGTGGCGAGCTGCT 73640
QY 1212 GTGCTGGCGGTGTGTGCTGGGACTGGGCAAGACCTCAAGCGCTCGCCACCGGC---C 1268
DB 73639 GCGTTCCTCGCGCGCGCAGACGCTGCGCCCGCAGGTGCTCGATCTCGCGATCGCTTTC 73580
QY 1269 CACAATGGAGGAGCAGCGCGCTGTGCGGACAGCAACATGCTCATCCACTTCGTGAAGAA 1328
DB 73579 CGACCTCACCATGCTGTGCGCGGCTGTATCGCGAGAGGTCAAGTCAACCTGATCCA 73520
QY 1329 CGGCGCCAAAGCGCTGTGTGTGCTGTTCGTAAGTTGTAAGCTGTGTGTCTTCAACAA 1388
DB 73519 CGGCGCGACCTTCTGGCGGTGAAGTGTGACGCTCTCCAGTTTCGAGCAGGTGATCGTCAA 73460
QY 1389 GCGCGTGTGTGTGCTGCGCGCGCGCTGCGGCAAGCAGTACGCTGTATCAAGCGGA 1448
DB 73459 TCTCGCGGTGAACGCGCCCGACGCCATGCCCGCGCGCAAGCTGATCATCCGACCCG 73400

misc_feature
/locus tag="RPA0632"
/notes="4 probable transmembrane helices predicted
byTMHMM2.0"
complement (2036..2377)
/gene="rnpA"
/locus tag="RPA0633"
complement (2036..2377)
/gene="rnpA"
/locus tag="RPA0633"
/function="InterPro IPR000100"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="probable ribonuclease p protein component
(protein c5)"
/protein_id="CAE26077.1"
/db_xref="GI:39647557"
/db_xref="GOA:Q6NC40"
/db_xref="InterPro:IPR000100"
/db_xref="UniProt/TREMBL:O6NC40"
/translation="MERLRADPIAVAGGLRINGPAFLVQLSRREDQGPVIRGVFTVT
KKGITATERNIRRLRELKTRDGRGLFPHSDYVLVGRRAALSRDFAVMGLDRLAL
ERVARSPPA"
complement (2401..2535)
/gene="rnpH, L34"
/locus tag="RPA0634"
complement (2401..2535)
/gene="rnpH, L34"
/locus tag="RPA0634"
/function="InterPro IPR000271 COGs COG0230"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="possible ribosomal protein L34"
/protein_id="CAE26078.1"
/db_xref="GI:39647558"
/db_xref="GOA:Q6NC39"
/db_xref="InterPro:IPR000271"
/db_xref="UniProt/TREMBL:O6NC39"
/translation="MKRTYQPSKLVKRRRHGFARLATTGGKVLAAARRARGKRLSA"
2866..4344
/locus tag="RPA0635"
2866..4344
/locus tag="RPA0635"
/function="InterPro
IPR003594:IPR003660:IPR003661:IPR004358:IPR005467 COGs
COG0642"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="sensor histidine kinase"
/protein_id="CAE26079.1"
/db_xref="GI:39647559"
/db_xref="GOA:Q6NC38"
/db_xref="InterPro:IPR003594"
/db_xref="InterPro:IPR003660"
/db_xref="InterPro:IPR003661"
/db_xref="InterPro:IPR004358"
/db_xref="InterPro:IPR005467"
/db_xref="UniProt/TREMBL:O6NC38"
/translation="MOAADKOPETHAWRGLEPKIGLSKLLLTITPLIALIILYV
PSIANYRLNRLDLAAANTALVLDASGRVPSDLSAQILDSIEARVAIVMGQOR
RLAASNLPTKIDVDMNRVTPWGAILDSFVWMLSEGNPAIRVVGPAAGNAEFIEV
IDAAPLRNARYFSSNLLVSLTITSLGLIYLALHFLFVRMRRLTAMVNFRRDP
ESPASIVVPGPDEITGLAERLSDMDRLVSLHOKSLAALGLAVSKINDRLNL
ASSQLSDLSVPDRVPAPKLVSLERLAFQCSLSYGRAQEAAPDRRLSILIE
PLVNDVRETAGADEASITWVSAIERGLTIDADPDOLFVLLNLVNAQAQLEGOPRN
DANLOQIRITGRBSGVTVLESDTGPVPQKARDHLFEAFHGSVRAGSGGLGLAIAA
ELVRHAGGBINLVEGTIGATFRISIPDPVLDLSLKRARA"

misc_feature
order (2944..3012,3394..3462)
/locus tag="RPA0635"
/notes="2 probable transmembrane helices predicted
byTMHMM2.0"
4594..4670
/gene="trna-Arg1"
/locus tag="RNA_7"
4594..4670
/gene="trna-Arg1"
/locus tag="RNA_7"
/product="trna-Arg"
/evidence=not experimental
complement (4784..5293)
/locus tag="RPA0636"
complement (4784..5293)
/locus tag="RPA0636"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="CAE26080.1"
/db_xref="GI:39647560"
/db_xref="UniProt/TREMBL:O6NC37"
/translation="MLKNLILFAANDHRPALAALLREHNPGTLVSDDLLELLEOLF
PRELQSRILICFGLPLTLQSDVVKQIGHGAYQFYAPLQYPLGLASTAQETDDGPCV
SVIAQSIGDAAPFDQIIDITVTLPANIESAPORDLIAFTRLARFWRLSHAIACETNP
AEPSTRLN"
complement (5320..6648)
/locus tag="RPA0637"
complement (5320..6648)
/locus tag="RPA0637"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="CAE26081.1"
/db_xref="GI:39647561"
/db_xref="UniProt/TREMBL:O6NC36"
/translation="MIPTSPSAKTELQASGLRLGTSQRRRLDLWLDRVGPVRLQRP
GEAPORDAGLIVTEPMGAGAEFLCNALPGCVVLPFGENDPDLKSLKTFEGTV
GPCAGGPHMVGWGDATLVADAGETARPRIVSCYOHGCDLWPLRQAASFCFL
AHIGPLMSECDRLVLAKEAFILQAWELPEPLIYIENGALLOAPLLPAQLDCDV
ALHWRNLSARTLYFGSDAAKALLGLRRRLAKHPAVWDGLDQWLSALSASA
LDTWPLRSYTLHGDPAPQSAVIVHDLPTGADLPDFGFASLVLTARRAGTGGRD
ALVMTSHQASGDGVAVILRLDITADSRVAVTTVETATSAFADNCGFGLLEMLCAW
RDEVDFAREAAASARLVAEIVFGQPIPHDFEARLDRQDAGAPSLIARG"

misc_feature
complement (6725..9127)
/locus tag="RPA0638"
complement (6725..9127)
/locus tag="RPA0638"
/function="InterPro IPR001951 COGs COG1011"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAE26082.1"
/db_xref="GI:39647562"
/db_xref="InterPro:IPR001951"
/db_xref="UniProt/TREMBL:O6NC35"
/translation="MSFVSIHTTPRAQNRANSFDVDTFLIRACTTPTGVPFLTYQL
SRVSDLCNMSEHFVQHRLAESRVREASRGATEINIDQIYARFPFLFGLERED
LNHLAEFSAELELCRINPDMLEQYRARRRAGNVGFI SDTYWTERLGRLLRACSP
GLTWFLYASCDHSGSKGDELFATYLRQQGVDASSAYHVGDNEDHADIRGAHRIIPR
HYQAPRLTTLRLLEDLLQQLMFKGPQTRFLDRGARTLRMVAARSAEQSAHHLGST
VVGVLAAQDPQVATRRADLQDGRRAVIGFLGRDGLSHRLWQQSRGPPAAVLEINR
RVSLASADTIQICELIRKIPRIDAAKLQDMKLPLPALADFPACDGGIATGABELA
DALPNLIEPHEIAMELAAGURGLDITRQIIPDDICTDULLVLDLGSVQALSRV

Query Match 3.9%; Score 76.8; DB 1; Length 349260;
Best Local Similarity 50.6%; Pred. No. 1.5;
Matches 239; Conservative 0; Mismatches 227; Indels 6; Gaps 2;
QY 819 CTTCTGTACACGACGCTCTGGAGAGACCCCGAGCGGATATGGAGGTATGGAGATCA 878

LOCUS AY386263 137241 bp DNA linear VRL 20-JAN-2004
DEFINITION Orf virus strain OV-IA82, complete genome.
ACCESSION AY386263
VERSION AY386263.1 GI:40019122
KEYWORDS
SOURCE Orf virus
ORGANISM Orf virus
VIRUSES; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae; Parapoxvirus.
REFERENCE 1 (bases 1 to 137241)
AUTHORS Delhon,G., Tulman,E.R., Afonso,C.L., Lu,Z., de la Concha-Bermejillo,A., Lehmkuhl,H.D., Piccone,M.E., Kutish,G.F. and Rock,D.L.
Genomes of the Parapoxviruses Orf virus and Bovine Papular Stomatitis Virus
J. Virol. 78 (1), 168-177 (2004)
REFERENCE 2 (bases 1 to 137241)
AUTHORS Kutish,G.F.
Direct Submission
Submitted (08-SEP-2003) African Swine Fever Research, Plum Island Animal Disease Center, U.S. Department of Agriculture, Agricultural Research Service, P.O. Box 848, Greenport, NY 11944-0848, USA
FEATURES
source
1. .137241
/organism="Orf virus"
/mol_type="genomic DNA"
/strain="OV-IA82"
/isolate="ORPA"
/isolation_source="nasal secretions of a lamb at the Iowa Ram test Station during an outbreak in 1982, then passaged in ovine fetal turbinate cells"
/db_xref="taxon:10258"
/country="USA: Iowa"
complement (2409..2858)
/codon_start=1
/evidence=not_experimental
/product="ORF001 hypothetical protein"
/protein_id="AAR98099.1"
/db_xref="GI:41018491"
/translation="MHPSRRLLGALLVALGFLGLFRPAAPPLPAALVBAQPVRA NGASVTCITVGDCGHMAVAHGGTLPVPLAGMHAFTASLRKGLLLNVATVH IYDRLAPFELTCAVAGGYNAAWATRAEWRRLQAQHRSEL"
complement (3002..3355)
/codon_start=1
/evidence=not_experimental
/product="ORF002 hypothetical protein"
/protein_id="AAR98100.1"
/db_xref="GI:41018492"
/translation="MTPTSRSLTTLVHLVRAARRPSPALVRALLAAGAAAGPDA AGRPLAALLRAALRRGGPLDARLLPILLARRGAAASARARGGARSAAGLLLAAGARV RLALAAAGRAAG"
complement (3713..3928)
/codon_start=1
/evidence=not_experimental
/product="ORF005 hypothetical protein"
/protein_id="AAR98101.1"
/db_xref="GI:41018493"
/translation="MPGEGVHYSHSGTSGVASTPGGAVKRPSSGGGIGPLEKRM CVGGGGGSKSGGVNGVNGGLGKI"
complement (4012..4491)
/notes="similar to Vaccinia virus strain Copenhagen F2L"
/codon_start=1
/product="ORF007 dUTPase"
/protein_id="AAR98102.1"
/db_xref="GI:41018494"
/translation="MEFCHTETQVRLSQNATIPARGSPGAAGLDLCSAHDCTIPSH SRVPTDLLIPPSGCGRIAPRSLGAVKHFDVAGVIDEDYRGNVGVLNFGNS DFEVKKGRDIAOLICERISCPAQVEVCLNDTRDGSFGTGLSGCGGRDTAWTIS"
complement (4554..6104)
/notes="similar to Vaccinia virus strain Copenhagen B4R"
/codon_start=1
/product="ORF008 ankyrin repeat protein"

/protein_id="AAR98103.1"
/db_xref="GI:41018495"
/translation="MLSRSEVVVPHADLLFRYLESGQVLETVRALVATDADVNFRRGE HGRTPLHLCVHFARHEQCAIVRVLLEAGADYKADSCGRTPHLYVQDVGPEVVA LMLEAGADVCDASVVVCDVSLGFCGPDGAKEAVLARLLDAGRVNEGTVGT PLHVAKQWIREDLRLRLBERGANCDNGLTPLAALCGGVSALVDMMLRAG ADACVVDVGRGTLTLLHARTAKISGMVRLTLGLAVDPAADVACGNVLMHTYCSG ERGVDFVLERGLDLNLRNNLTALHRAVESHGACCLVRMGAELEHMSASGLCAV SMLENNRVATAAVLARPPPTELLARALITSEQWGHFRRSEALCLCQVELARGGV ARMLAERALADYATVIRACEQBIASMRVRCHTDATTLLDVLRAADAKALFVQNALFE RAEFPITVGTALFGKICMRLRVSLAEQIAGLMCPALPPEIVTSLCFLSYSLDL L RRAMLTRP"
complement (6265..7626)
/notes="similar to Vaccinia virus strain Copenhagen F11L and Molluscum contagiosum virus MC018L"
/codon_start=1
/evidence=not_experimental
/product="ORF009 hypothetical protein"
/protein_id="AAR98104.1"
/db_xref="GI:41018496"
/translation="MFFWFCGRPRMSALARLFGRRGRNOCVPTAGCMPLRRPRDG ADGGSDDDYGYGTDTSVASSVAPVEARPLVLTVPGTCPVLVDSVFTRIENTSVI VTHGSEWTDLSLFESETHRAPARICKSEPSRVVDANMHFEVGSGLKIKHDEWVR HMAPTVAAYVTIICVKEGMAGISVNTKELKTNMQEGDVLVVPARSVEPLPQVGGI DAYIILVPTAEICDLGFEPFSPSIDQEAQVRVANVAETRRKACKIVNQIVKLAVD LEKCYEICNVIMITEFNVDYDGAQMTASVDAIARGMGAQTALCKRSLRGR PLKSAALTRNHSILEHITSLFAVFAQTPEGTTPHDPDLMARMERDULRKSSELDWD ILQLASDLSDTLPRCECIERYIHHMVCVSEAGVGRNELVQRLTMIAGAGVRLTSGSGG VPV"
complement (7651..9582)
/notes="similar to Vaccinia virus strain Copenhagen F12L and Molluscum contagiosum virus MC019L"
/codon_start=1
/product="ORF010 putative EEV maturation protein"
/protein_id="AAR98105.1"
/db_xref="GI:41018497"
/translation="MLGFWKVCVGPDRQWDEPPTDLVPGEYALAKARDGKTMIF DAGARAAAVPLSLKSVRISCTFIDARADALAPPAPATQRFATVSRDDPCYITPDS SSPFLDIILQRAETPELRAEFEDPPPRDALSMDELQNLSDVGLYDLRMVSHDAVR RIGTLPRQLTLLDMNVCCGVSYAIWIKDPTSYRPTDEVSHDVYTAASAWQAPH VHAQNSVAAVVFVRFGSDGVRLNSIATNAGPVPFSGCAKLIVRDLFWMLTTCKEN ITVGYFSFEDPTROVMELAGKGTRENTLLSLRGNRVAFLHRAAPATLFSYVC EFWGCVVDVEDVVSASAESEAAAAAQAALSDAALAHQAALARIFPSCDMAAF RGLREMLVNAARGCAACPVHFSALNVEAALFEBETGASVPADARCFRLPAREV ISATQLYGVKPRVYSLKTRGLLAIALCEVTRSAEVKIPVLPFPADESLRSLVLTSDV IETALLKGYEIRVVCALEWGRCDVFLKAGLEAQTATRELNIPTSLMSRTASMP L ALETDDSPSGRCSAVVRAFAASYARRAVHGFIERVDCHATAAFVLRHGVDYRFWVRERA ARCLAGSGIQSLPSVR"
complement (9653..10789)
/notes="similar to Vaccinia virus strain Copenhagen F13L and Molluscum contagiosum virus MC021L"
/codon_start=1
/product="ORF011 putative EEV envelope phospholipase"
/protein_id="AAR98106.1"
/db_xref="GI:41018498"
/translation="MWPFSIPLGADCRVETLPAPVAVSLAQNMSTLDCFTAITESA KFLYICFCNLSSTKGVGVKDLAKETLAKGVNVTLVVDVQSKDADELRAGVY YKVKVSTREGVNLGSLFWSLDAGHWYVGSASLTGSGSVSTIKNLGLYSTNKLAWDL MNRVNTFVMIVEPKVPFTRLCVAVVTPATNFHLDHSGGVGFFSPRFLGFGVRLT LKWRVHLRIENAKNSIDLSLSNVPVKHASAVEWPRIIDALLRAIDRGVVRVIL TEWKVADPLVSAARSLDDFGVSDMSVKFVPGRDDAANKLLI VDDTPAHUTV ANLDGTHRYHAFVSNABKGDIVKDLSAVFERDWRSEFCKPIN"
complement (110818..111087)
/codon_start=1
/evidence=not_experimental
/product="ORF012 hypothetical protein"
/protein_id="AAR98107.1"
/db_xref="GI:41018499"
/translation="MLIYGLPRLEVEDLIDRMDDVQVSGDLMPPTDPVPEDVAQCSPE DMANMLDGLIEDIQLKSELQTLTSEMEBELLREBELARLLKDFP"
complement (11396..11635)
/codon_start=1
/evidence=not_experimental

```
/product="ORF013 hypothetical protein"
/protein_id="AAB98108.1"
/db_xref="GI:41018500"
/translation="MGVWMCGLFVWAGWIARRVRFVVALRORVSRDKGYVAVIQTCD
DDYFTFEFFDDQKQVALLRDVSRVVAAPAGVTE"
complement(11671..11952)
/notes="similar to Molluscum contagiosum virus MC026L"
/codon_start=1
/product="ORF014 modified RING finger protein"
/protein_id="AAB98109.1"
/db_xref="GI:41018501"
/translation="MPVKVKQMLAVSCTILACSDVCYIKNRSPGICINACPGACPF
QMVVCGHYVHCITPTNTEVCFVRSALQKAEPLDMSTSKSAREFIV"
complement(11955..13574)
/notes="similar to Molluscum contagiosum virus MC027L"
/codon_start=1
/evidence=not experimental
/product="ORF015 hypothetical protein"
/protein_id="AAB98110.1"
/db_xref="GI:41018502"
/translation="MEBELRLIAQAMMDLSVAHISRLRLTLKLVNTFTAVACLR
EALARKSWPPGRLGDIHEDMTVLEYVMQVEVENARENTEPYLCVIOAHNADVLDD
CLREADMEYGRIGRVECRPHLVADRCSCLSAFSEDACVFTNLMRQS
RTMARCIVRAHSGGSDYDELVDVHRQNVKRLWPRVVANVEDVIAADLEGLMAY
CRTCLMSARVASGGLDILKDRMPILIVECIEIHAESNLDALCLVNVVDVLTGTHG
LHARLERLSGKALYGANVAAAPCDTMDLLSYLERLPRELAAFAVAFAIVKSVYT
GVNLEKLSLKWLELPIVPMASALNDLVRLMSGRLLPLLPPOPMEDHGESGHH
DPMVLMVPVNVYTFVPLAAMPPELSRGAVESALARADTAYDVAPLAAGRAELT
IPAAGREQBEVICSSAHLVCIVLMDKAGTGTVEADAARHLGNLDAEMTLTSLCRE
KIAKRTVDGKRFALNPLEARDSPWVFN"
complement(13644..14423)
/notes="similar to Vaccinia virus strain Copenhagen F16L
and Molluscum contagiosum virus MC029L"
/codon_start=1
/evidence=not experimental
/product="ORF016 hypothetical protein"
/protein_id="AAB98111.1"
/db_xref="GI:41018503"
/translation="MIRIDGMEVEVDENATRAARGAMLLSTSFMSMDVIVQRELA
RQCYKNGMTPCVSVTRIGRIDASNGEDTWEARSHDVRHVVLFKASVVDHVDLFL
VGLGELGLDLYFVADGFPVDAFPMKFPDGMETMLRCVLRDVEDVGGGLLHPR
EPGASQCLMDRYMFPFGLLAMOAHFARRSRRAPTPLPRASPRPQTRRSPLRGVAD
SAMYQALADARANNQLRRRRRDDVLSICRLNDNMRVDD"
14677..14994
/notes="similar to Vaccinia virus strain Copenhagen F17R
and Molluscum contagiosum virus MC030R"
/codon_start=1
/product="ORF017 DNA-binding phosphoprotein"
/protein_id="AAB98112.1"
/db_xref="GI:41018504"
/translation="MDGGVHKRPPIVNDGNGKVLVLYLRMCVEPFAKCEGRASCV
LKMDDPRSPSCERRSLPFPSPCPMTTPGSPLOAPLMRTQMLQGLFDAKNGEOMC
RRQ"
complement(15018..16436)

Query Match 3.98; Score 75.6; DB 13; Length 137241;
Best Local Similarity 43.18; Pred. No. 2.6;
Matches 463; Conservative 0; Mismatches 609; Indels 2; Gaps 2;

QY 618 CGACCAATTTGACATCGGCGCCGAGCGCGCTACCTGAGCAGAAAGCTGGAGCGCGT 677
DB 20215 CGAGTCTTTCAGCGCCGTGTCAAGCGCGCTTCGACGCGCGCAGCTCTTTGCGGGC 20274
QY 678 GTGGAGCAGAACACCCAGGGTTTCGATCCCTACGTGCCGCTGGCTGCGCGCCCTACTA 737
DB 20275 GCTCTCGCGGCACTTCGAGGCTTCGTTCGCGCATCCGGGCTTACCGCGGCACTTCGC 20334
QY 738 CGTGTGATTTGGCGCTGCCAGCTTTGGCCACCGCTTCGACGAGGAGCGCTGGAGCG 797
DB 20335 GGAATCTGTGCCGACGCGCTTTCATGAGAGGTGGCGCGCGCGCGAGCTTCGCGGA 20394
QY 798 GCGCCCATGTTCCGCGCCACCTTCCTGTACACGAGTCTGGGAGGACCCCGAGCGGGA 857
```

```
DB 20395 GGTCAATTGGCTGAACACGACATCCCGCAACCCGCTCTTCGCGAGCGGAGCCCGT 20454
QY 858 TATGAGGTGATGAGATCAACCCCAAGACACGGTGTGCTGACCTGACTAGCCGCGGCTG 917
DB 20455 GCGCAGACGGGAGCTCATCTTCGAAAGACCTTTCGCAAGACCGAGTTCGCGCCGCTCAA 20514
QY 918 CAATGCCCTGAACCTCTGCTGTCAGGGGCGCGCCAGGTGGTTCGCTGAGTGCACACCC 977
DB 20515 GCGCTTCGCGCTGTCGGCTGCTGCTGCGGCTTCCTCGTGAAGAGGACCTTGGCGG 20574
QY 978 CGCGAGTCGGCGCTTCTGAGAGCTGAAGAGTGGCCATTCAGCAGCTGAGTGTGAGGA 1037
DB 20575 CGAGTACGCGGACCAACGACCGCAGACCTGTTTACGCTGCTCAGAAAGCCGCGGGCC 20634
QY 1038 CGTGTGGAGCTGTTTCGGCGAGGGCTGCACCGCGCATTTGAGAGCTGTACGAGAGAA 1097
DB 20635 CGTGGCGCCA-CAGCGCGCTACGAGAGCATTCGCGAGTACCTTCTCCCGGAGACAGGC 20693
QY 1098 GCTGGCGCCCTTCTGCTGCAACACCA-GCCACAACTTTCTGGTCCAAAGCGCTCTGTACT 1156
DB 20694 CAGCACCCTGGGTCTGGCTGACGCGCGCTGGCCGACGCGGAGTGTACCGGACC 20753
QY 1157 TCAGACACCGCTGTACTACAGGGCGGCATGGGCAAGTGTGCTGGGTGCTGAGTGCC 1216
DB 20754 GGCCCGCGGCGACGCTCTACGAGCGGTGCTCAGCTACGCGTACTCAGAGGTCAAGCAG 20813
QY 1217 TGCCCTGTGTGGAGCTGGGCAAGACGCTCAAGCGCTTCGCCAACGCGCCACATGG 1276
DB 20814 GCGCGCTGAACGCGCAACACGCTCGTGTGATCCGCGTTCGAGGACGACCCGACATCA 20873
QY 1277 AGGACGAGCGCGCTCTGTGGGACAGCAACATGTCTCATCTTCTGTAAGAAAGCGGCCA 1336
DB 20874 AGGCTCTGCTGTCGAGCTCATCTACGCTGCGCGGACATCTCGGCGTCTGTGACT 20933
QY 1337 AGCCGTGTGTGGCTTGTTCGTAAGTTCGTGAGCTGTGTCTTCAACAAGCGCGTGC 1396
DB 20934 CCGCAACGAGGAGTGGCGGAGCTACTTCTGTAGTCTGTACCGCGAGAACTTCGTGACG 20993
QY 1397 TGTGTTTCGGCGCGCGTTCGGGCAAGCAGTACGCGTGTATCAAGCGGAGCGGATCC 1456
DB 20994 GACGCACTTTCACCTCGGACGCGCTTCGCGACGACCTTCTTCGCGTGTTCGCGCGC 21053
QY 1457 CCATTGAGAACTACATCTCGCGCACCATGGAACGCGCTGGCGGAGAACTCCGACGTGCGCA 1516
DB 21054 TCAGCGCGCTTCTTCGAGCGCGGCGCATCCGCGAGGCTTCAGTGCAGACGCGGCG 21113
QY 1517 AGCAGAACTACTTCTACTACACTGCTCACCGGCAAGTTCCTGCGGCGACAACTGCGCCA 1576
DB 21114 TGGAGAGCGCTTTCAGGACATGGACCTCAACAAACGCTTTCATGTCGACCTCATCTACG 21173
QY 1577 CTTACTTCGCGAGGCGGCTTCGCCACCTCTCAAGAGTGCCTGAGTGCAGAACTGACCG 1636
DB 21174 ACTCCGTGAGCCCGAGCTGCGCGCGCGAGCGCGGCTTCGCGCTGCGGCTGCACAAACG 21233
QY 1637 TCTCCACCACTTCTTCATGGAGGAGCTCAAAGCGGCGACCTACACCAAGGTGA 1690
DB 21234 AGACTCCGACTACTTCATCCGAGGAGTACAACACTACCTCTTCTCTCAGCGAGA 21287

RESULT 13
AP006840_19/c
WPCOMMENT
Sequence split into 36 fragments LOCUS AP006840 Accession AP006840
Fragment Name Begin End
AP006840_00 1 110000
AP006840_01 100001 210000
AP006840_02 200001 310000
AP006840_03 300001 410000
AP006840_04 400001 510000
AP006840_05 500001 610000
AP006840_06 600001 710000
AP006840_07 700001 810000
AP006840_08 800001 910000
```

1258 GCCAACGCCCCACATGAGGAGGAGCAGCGCGCTCTGTGGGACAGCAACATGCTCATCC 1315

Db 53345 GCGCACATGAGCGGATGAGCTGATCGCCAGGTGGAGCGCGCTGACCTGGACAGCC 53288

RESULT 14

LOCUS BX842626 75337 bp DNA linear PLN 17-APR-2005

DEFINITION Neurospora crassa DNA linkage group I BAC clone B18P24.

ACCESSION BX842626

VERSION BX842626.1 GI:38566958

KEYWORDS

SOURCE Neurospora crassa

ORGANISM Neurospora crassa

REFERENCE 1 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.

AUTHORS Schulte, U., Aign, V., Hoheisel, J., Brandt, P., Fartmann, B., Holland, R., Nyakatura, G., Mewes, H.W. and Mannhaupt, G.

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 75337)

AUTHORS German Neurospora genome project.

TITLE Direct Submission

JOURNAL Submitted (26-NOV-2003) MIPS, Institut fuer Bioinformatik, GSF-Forschungszentrum fuer Umwelt und Gesundheit, GmbH, Ingolstaedter Landstrasse 1, D-85764 Neuherberg, FRG, E-mail: G.Mannhaupt@gsf.de Project Coordinator: Ulrich Schulte, Institute of Biochemistry, Heinrich-Heine-University, D-40225 Duesseldorf, E-mail: ulrich-schulte@uni-duesseldorf.de

COMMENT BAC clone 18P24 (strain OR74A) is available at the Fungal Genetic Stock Center, <http://www.fgsc.net>

Sequencing was performed by MWG Biotech AG, Ebersberg, Germany, <http://www.mwgdna.com>

Information on performance of analysis and a more detailed annotation of this entry and other sequences can be viewed at: <http://mips.gsf.de/proj/neurospora>.

FEATURES

source 1. .75337

/organism="Neurospora crassa"

/mol_type="genomic DNA"

/db_xref="taxon:5141"

/chromosome="I"

repeat_region 2373..3008

/note="36 bp gaagaaagaagaagaagaaga tandem repeat"

gene 4630..5673

/gene="B18P24.020"

gene complement(4630..5673)

/gene="B18P24.020"

CDS complement(4630..5673)

/gene="B18P24.020"

/note="contains EST gb:BG278910, BG279921, BG278901, BG279679, BG279920, BG278902, BG278911, BG279791"

/codon_start=1

/product="putative protein"

/protein_id="CAE76260.1"

/db_xref="GI:38566959"

/db_xref="UniProt/TREMBL:O7SDH3"

/translation="MSTKPLIKLSTPTANFNPIPEPTDLRSAVTLPSALSPTPLS ALRSYAFRDLPSALPSAGLPGSPSATLKVDHDLQKPTTPPPVAYLFLKCMQLVSLPSLPGPTQTKVSLHRTSPASLSYSTSHSESSVDATTDDSKBKEGQLESAPPTARTDVSSEPEIKVEAEEDKKDKEDKEDVDNKEDSKSPAPITISQPTPTQSLVPMSPATGCAAVFPPTKLPASPAITSALYSPTPLSAASVSPSPDWEALTKLRRAVYSLKRSAPSADAPAVSTPTSPVFAKESRSSIRHREVTRTVTYTPRMAPAPKGRKDKIDATPAKS"

exon complement(4630..5673)

/gene="B18P24.020"

/number=1

repeat_region 7464..7512

/note="49 bp poly(a) tandem repeat"

gene 14161..15111

/gene="B18P24.030"

gene complement(join(14161..14342,14412..14477,14566..14717),14924..15111)

```
/gene="B18P24.030"
complement(join(14161..14342,14412..14477,14566..14717,
14924..15111))
/gene="B18P24.030"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAE76261.1"
/db_xref="GI:38566960"
/db_xref="UniProt/TREMBL:Q7SDH4"
/translation="MARSAAPRGPGMGEAARPHREFPQPPPAQWKRCGPOOQTSF
RNDPAGLPNHPAPVAGSTRGDDAVLAKODAAQSGIVSGRCPLGVFVPCAGN
ESDMWNEAVRYNLPIARVKVGLDMSLVSPDLAVLDHPCRPVLAGFGLRCD
GGGSRWGTMLVLRYLRAAALRAALRMSLSV"
complement(14161..14342)
/gene="B18P24.030"
/number=1
complement(14343..14411)
/gene="B18P24.030"
/number=1
complement(14412..14477)
/gene="B18P24.030"
/number=2
complement(14478..14565)
/gene="B18P24.030"
/number=2
complement(14566..14717)
/gene="B18P24.030"
/number=3
complement(14718..14923)
/gene="B18P24.030"
/number=3
complement(14924..15111)
/gene="B18P24.030"
/number=4
1595..17623
/notes="29 bp poly(g) tandem repeat"
19883..20827
/gene="B18P24.040"
19883..20827
/gene="B18P24.040"
/notes="similarity to HAPB protein, Aspergillus nidulans,
TREMBL:ENHAPB 1
contains EST gb:AA898026, AA898402"
/codon_start=1
/product="related to CCAAT-binding factor HAPB protein"
/protein_id="CAE76262.1"
/db_xref="GI:38566961"
/db_xref="GOA:Q7SGY4"
/db_xref="InterPro:IPR001289"
/db_xref="UniProt/TREMBL:Q7SGY4"
/translation="MMEYAOYQPAQNSHNPMTAYSSAGGNSITSPSSOHLPTS
PILOTQHQSPSTQGHSMYQTVGVQONHMGMPLOQAALANTAASGTNYNPMYA
ADNLQSPMSGVNSKQVRAGRFPQNNISPPGQRLSQVASPGVNAQAWLNHA
SRGAVPPMTPAQMPQPPQSPENASGASVPLYNAKQPHRLKRVARQKLEQRL
LTNKRKPYLHESGRHNMRRPRGPGRFLTADAVQMRDKVNGDVQDGSQSSVT
AGSKTTGGTKRKAEISTGAPNPKAKAAPESPEDDASD"
19883..20827
/gene="B18P24.040"
/number=1
21444..21471
/notes="28 bp gcacgcacgcacgcac tandem repeat"
23226..26083
/gene="B18P24.050"
join(23226..23780,23851..24177,24239..26083)
/notes="no similarity to N-terminal half of the protein
similarity to S-adenosylmethionine:diacylglycerol
3-amino-3-carboxypropyl transferase btaA, Rhodobacter
sphaeroides, TREMBL:AF329857.1
contains EST gb:BF072578, BF072579"
/codon_start=1
/product="related to S-adenosylmethionine:diacylglycerol
3-amino-3-carboxypropyl transferase btaA"
```

```
/protein_id="CAE76263.1"
/db_xref="GI:38566962"
/db_xref="GOA:Q7SGY5"
/db_xref="InterPro:IPR000051"
/db_xref="UniProt/TREMBL:Q7SGY5"
/translation="MGDNSAMASHGHMGNIISLPSVSTRTLADLNFNNTTYSIFFT
GVAVLFLVTTSNYSRKTTKNEDDNEGNPSLSKSLLLFCYSCFIKPHATAGTTQTQ
QDALLSFYSQADIYDATRTGLTKGREDMALAAQRLRYKVEAGLGGAGDGLSKR
QSNKTCVTAGTGTGTRKPIWVDVGGGTGWNIEAMAKFVNVSEFPKTVYLVDFGSP
LCEVAKTRFARLGWENVVICDARKFRLEDVEDVDESGSDSPSSGMMGETKP
GRHAGNELTMSYSLSMPPDYFSLIDSLESLAPHLIAVDVFAQSKVDFTPRNTG
GLMNRHVGTFARNFWRFDADKVSLEPPARPDYLEYRGTVTIYNARNNTILGAIPYI
WIGCKKPPSTSLPHEIIVEHIDAIATSPSPRLVGRHSSATNALAPAVGRTAPE
MRSAFNATAIENISANPLPFSFYQNHWRIYYDDQLPHTQFNDEYIYIAFTWEDSRV
DRELLNLGDDVLAITSAGDNILSYLMQSPARVHAIDLNPQNHLLEKLVASFTFLD
YDVMKI FGEKGHPDPSLLISKLSPHSGRAFOYLSNAHIFTDPAGRLVDTGGR
YAIRFRWISTLTFCCSAVRRLSTPTLEGQSIYHTKIRPCLLRFNGLVLSDAF
LWSALGVPKNQVMAEADYHRSISSTTTPSKRPSRAEILUHTYTSHLDVLSH
GSDNPFYLVCLVGTQYTPQCHPDYLSAAHSILSAPGDFGLRIHTDEIOELVARQP
GLTIVAVMDSMDWFDPPSPSEKGRGKAREQVRLNRLALKVGKVLRSAGVEPWY
VRVFEEGFGARRVGCRESGRGDQECIDRVNMYASCWILEKMELEELVDSA"
23226..23780
/gene="B18P24.050"
/number=1
23781..23850
/gene="B18P24.050"
/number=1
23851..24177
/gene="B18P24.050"
/number=2
24178..24238
/gene="B18P24.050"
/number=2
24239..26083
/gene="B18P24.050"
/number=3
27766..28499
/gene="B18P24.060"
complement(join(27766..28118,28258..28275,28370..28499))
/gene="B18P24.060"
complement(join(27766..28118,28258..28275,28370..28499))
/gene="B18P24.060"
/notes="match of 21 ESTs
similarity to succinate dehydrogenase precursor (EC
1.3.99.1), YDR178W, Saccharomyces cerevisiae, PIR:A54380
contains EST gb:A1397618, AW18559"
/codon_start=1
/product="related to succinate dehydrogenase precursor"
/protein_id="CAE76264.1"
/db_xref="GI:38566963"
/db_xref="GOA:Q7SGY6"
/db_xref="InterPro:IPR007992"
/db_xref="UniProt/Swiss-Prot:Q7SGY6"
/translation="MASVARSSALLKQVAAQSSVAANGLRVAAFTHTSRKSLPPPPQ
RIEGTVNDVEVPVPPSPSGSYHTFDVVAAGLILTVAPFAAGSLNPTMDVLAAT
ILHSHTKFGNIIVDYVPSKRVPKRVETWGLNAATLVGLALYEFETDVLGTETI
KRVWKA"
complement(27766..28118)
/gene="B18P24.060"
/number=1
complement(28119..28257)
Query Match 3.7%; Score 73; DB 15; Length 75337;
Best Local Similarity 51.0%; Pred. No. 6.2; Indels 0; Gaps 0;
Matches 172; Conservative 0; Mismatches 165;
QY 779 ACCGAGGCGCGTGTGAGCGCGCCCATGTTCGCCGCCACCTTCCTGTACACGAGTCGT 838
|||||
DB 24804 ACCGAGTCAACTCCCGAAGCACACCCAGTTCATGACGAGTACATCTACGCCTTTACCT 24863
|||||
QY 839 GGGAGGACCCCGAGCGGATATGAGGTGATGAGATCAACCCAGGACACCGTGTGTA 898
|||||
DB 24864 GGGAGACTCGCGCTGCGACAGAGAACTCCTTAACCTCGGGCCGCGACGCTGCTAG 24923
|||||
```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 13, 2006, 22:58:40 ; Search time 367 Seconds
(without alignments)
9430.279 Million cell updates/sec

Title: US-10-620-914-44

Perfect score: 1947

Sequence: 1 atggggctcggtcggtgacgg.....gogccaaagagacaactaa 1947

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1303057 segs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:*

- 1: /cgn2_6/prodata/1/ina/1/1.COMB.seq.*
- 2: /cgn2_6/prodata/1/ina/5.COMB.seq.*
- 3: /cgn2_6/prodata/1/ina/6A.COMB.seq.*
- 4: /cgn2_6/prodata/1/ina/6B.COMB.seq.*
- 5: /cgn2_6/prodata/1/ina/H.COMB.seq.*
- 6: /cgn2_6/prodata/1/ina/PTUS.COMB.seq.*
- 7: /cgn2_6/prodata/1/ina/PP.COMB.seq.*
- 8: /cgn2_6/prodata/1/ina/RE.COMB.seq.*
- 9: /cgn2_6/prodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	69.4	3.6	3003	3	US-09-252-991A-14790
2	69.4	3.6	3132	3	US-09-252-991A-14540
3	69.4	3.6	3411	3	US-09-252-991A-14823
4	68.2	3.5	2830	2	US-09-010-928B-1
5	63	3.2	1631	3	US-09-302-540-6564
6	63	3.2	2034	3	US-09-302-540-8514
7	63	3.2	2730	3	US-09-302-540-489
8	63	3.2	6855	3	US-09-302-540-897
9	62.4	3.2	1082	3	US-09-881-165-4
10	62.4	3.2	1167	3	US-09-302-540-2993
11	62.4	3.2	1250	3	US-09-302-540-362
12	62.2	3.2	3579	3	US-09-302-540-7318
13	62.2	3.2	5656	3	US-09-302-540-694
14	61.8	3.2	981	3	US-09-302-540-6237
15	61.8	3.2	13332	3	US-09-302-540-1047
16	61.8	3.2	19726	3	US-09-302-540-1164
17	61.6	3.2	1852	3	US-09-369-852-4
18	61.4	3.2	1281	3	US-09-302-540-7495
19	61.4	3.2	6269	3	US-09-302-540-726
20	60.8	3.1	951	3	US-09-252-991A-10513
21	60.8	3.1	1653	3	US-09-252-991A-10546
22	60.8	3.1	1701	3	US-09-252-991A-10316
23	60.4	3.1	1515	3	US-09-302-540-9184
24	60.4	3.1	9556	3	US-09-302-540-982

25	60.2	3.1	1563	3	US-09-302-540-8878	Sequence 8878, Ap
c 26	60.2	3.1	10528	3	US-09-302-540-945	Sequence 945, App
27	60	3.1	9818	3	US-09-302-540-987	Sequence 987, App
28	59.6	3.1	1941	3	US-09-302-540-6922	Sequence 6922, Ap
29	59.6	3.1	4359	3	US-09-302-540-594	Sequence 594, App
30	59.2	3.0	1773	3	US-09-302-540-7746	Sequence 7746, Ap
31	59.2	3.0	4854	3	US-09-302-540-768	Sequence 768, App
32	58.8	3.0	1293	3	US-09-302-540-6161	Sequence 6161, Ap
c 33	58.8	3.0	1675	3	US-09-302-540-302	Sequence 302, App
34	58.4	3.0	1011	3	US-09-252-991A-12311	Sequence 12311, A
35	58.4	3.0	1236	2	US-08-440-856A-1	Sequence 1, Appli
36	57.8	3.0	1326	3	US-09-302-540-5594	Sequence 5594, Ap
c 37	57.8	3.0	41927	3	US-09-302-540-1288	Sequence 1288, Ap
38	57.6	3.0	4345	3	US-09-949-016-2755	Sequence 2755, Ap
39	57.6	3.0	8346	3	US-09-949-016-14497	Sequence 14497, A
40	57.4	2.9	975	3	US-09-365-150-3	Sequence 3, Appli
41	57.4	2.9	2261	3	US-09-158-767-1	Sequence 1, Appli
42	57.4	2.9	2261	3	US-09-713-794-1	Sequence 1, Appli
c 43	57.4	2.9	9191	3	US-09-302-540-918	Sequence 918, App
44	57	2.9	1059	3	US-09-252-991A-7911	Sequence 7911, Ap
45	57	2.9	1338	3	US-09-252-991A-7764	Sequence 7764, Ap

ALIGNMENTS

RESULT 1

US-09-252-991A-14790
; Sequence 14790, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 14790
; LENGTH: 3003
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-14790

Query Match 3.6%; Score 69.4; DB 3; Length 3003;
Best Local Similarity 43.8%; Pred. No. 0.00015;
Matches 454; Conservative 0; Mismatches 571; Indels 12; Gaps 3;

QY	521	TGGGCGTTCGCGACTTCTACGTGAGCGGCAAGTACGACCTGCCCTGGCCGACATGCCT 580
DB	1965	TGGGCGTTCGCGCGACCAACCCGACCTCAACACCGCGCGCTGCTGATCAACTCAAGC 2024
QY	581	GGTTCGCGCGCTTCTCTCGCGATCGATCTTCGACATCGACACATTCGATCGGCGCCG 640
DB	2025	CGCACAGCGCGCGAGCTCACCGCAGCGAGTGATCCAGCGCTGCAGCGCCGACTCG 2084
QY	641	AGCGCGCGCGCTTACCTGGAGAGAGAGCTGGAGCGCTGTGGGAGAGAGAACCCAGGGTT 700
DB	2085	ACCACCTGCGCGGATCAAGCTGTATATGACGCGGTGCGGACCTGACCATCGAGACC 2144
QY	701	CGATCCCTTACGTGCGGTGGTGGCGGCGGCTTACTACTGTTGGATGGCGGCTGCCA 760
DB	2145	GGGTGCGCGCGACCCGACCTGACCTTTCAGGACCGCGGACCGCGAGCTGCTCGCG 2204
QY	761	GGGTGGCGCGCGCTTGCAGGAGCGGCTGGAGCGGCGGCGGCTGTTCCCGCCACCT 820
DB	2205	AGTGGGTGCGGAGCTGGTGGCGGCTGCGAGGATGTCGCGAGCTGCGGAGCTGCCA 2264
QY	821	TCTGTACGCGAGTCTGTTGGAGGAGCCCGGCGGATATGGAGGTGATGAGATCAACC 880

Db 2265 GCACGTGGCAGGCAAGGGCTTGCAGGCTTACCTGACATCGACCGCG-----ACACCG 2318
Qy 881 CCAAGGACACGGTGTGACCTGACCTAGCGGGCTGCAATGACCTGAACTGCTGTGTC 940
Db 2319 CTTGCGGCTTCCGGCTGAAAGCTCTCCGACATCGACAGCGTGTCTACACGCGCTTCGGCC 2378
Qy 941 AGGGGGCCGGCCAGGTGTGCTGGTGAACCTGCAACCCCGCGAGTCGGCGCTTCTGAGC 1000
Db 2379 AGCGCTGTATCTGACCATCTTCAACGAGCCACCCAGTACCGGTGTGCTGAGGTG 2438
Qy 1001 TGAAGAGGTGGCCATTCAGCAGCTGGAGTTTGAAGACGTGTGCGAGCTGTTCGGCGAGG 1060
Db 2439 CGCGCAGTTCCAGCTCGGCCCGCAGGCCCTTGGAGCAGCTCTAGTGCCTGCTCCAGCGACG 2498
Qy 1061 GCGTGACCCCGCGGATTTGAGAGCTGTACGAGAAGAGCTGGCGCCCTTCTGCTGCGAAA 1120
Db 2499 GCACCCAGGTGCGGCTTGTGAGCCTTGGCAAGGTGGAGAGCGGCATACCTGTGTCGCA 2558
Qy 1121 CCAGCCACACTTCTGCTCCAGCGCTCTGTACTTCCAGCAGCGGCTGTACTACCGAG 1180
Db 2559 TCAACCATATCGCCCAAGTTC---CCCTCGGCGACCTGTCTTCAACTGGCCAGGGTT 2615
Qy 1181 GCGCATGTGGCAAGCTGTGTGGGTGCTGCAGTGCCTTGGCGCTGTGCTGGGACTGGGCA 1240
Db 2616 ACTCCCTGGGAGGCGGTGAGGCGATCCGTGCGCTCGAGGCCAGCTTGGAGCTGCGC 2675
Qy 1241 AGACCGTCAAGCGCTCGCCACCGCCGCAATGAGAGGAGCAGCGCGCTGTGTGGGACA 1300
Db 2676 TGAGCATGCGAGGCGAGCTTCGCGGCGCGCGCTTGGCGCTTTCGAGGCCCTGCTGCAACA 2735
Qy 1301 GCAACATGCTCATCC---ACTTGTGAGAACGGCCGCGCCAGCGCTGTGCTGGCTGTG 1357
Db 2736 CGTGTGCTGTATCTCGCTCGGTGTGACCATGTACATGCTCCTGGGATCCTCTACG 2795
Qy 1358 TCAAGTTCTGTGAGCTGTGTCTTCAACAGGCGCTGTGTGTGCTGGCGGCGCGTGC 1417
Db 2796 AGAGTTTCATCCATCCGTTGACCATCTCTCGACCTGCTCGCGCGGGGTGCGCGCG 2855
Qy 1418 CGGCAAGCAGTACGCGCTGTATCAAGCGGAGCGCATCCCATTTGAGAACTACATCGCG 1477
Db 2856 TGCTGGCGCTGTATCTGCGCGGCGAGGAGATCGGCATCGTGGCGATCATCGGCATC 2915
Qy 1478 GCACCATGAGCGGTGCGGAGAACTCGCACGTGCGCAAGCAGAACTACTTCTACTACA 1537
Db 2916 TGCTGTGCGCATCTCAAGAAAGCGCATCATGATGATGATGCTTCCGCTCGACCGCG 2975
Qy 1538 ACTGCTCACCGGCAAG 1554
Db 2976 AGCGCAAGGAGCAAG 2992

RESULT 2

US-09-252-991A-14540
; Sequence 14540, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 14540
; LENGTH: 3132
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-14540

Query Match 3.6%; Score 69.4; DB 3; Length 3132;
Best Local Similarity 43.8%; Pred. No. 0.00015;
Matches 454; Conservative 0; Mismatches 571; Indels 12; Gaps 3;

Qy 521 TGGCGCTTGCAGCTTCTACGTGAGCGGCAAGTACGACCTGCGCCCTGCGCCAGATGCGCCT 580
Db 1799 TCGGCGTTCGACGCGCAGCAACCCGACCTCAACACCGCGCGCTGCTGATCAACCTCAAGC 1858
Qy 581 GGTGCGCGCTTCTTCTGGCGATCGATCTTCGACATCGACAACTGATGACATCGGCCCG 640
Db 1859 CGCACGCGAGCGCGAGCTCACCGCAGCGAAGTATCCAGCGCTTGCACCCGAACTCG 1918
Qy 641 AGCGCGCGCTTACCTTGGAGCAGAGCTGAGCGCGGTGTGGGAGCAGAAACACCCAGGGTT 700
Db 1919 ACACCTGCGCGGATCAAGCTGTATCATGACGCGGTGACAGGACCTGACCATCGAGGACC 1978
Qy 701 CGATCCCTACGTGCGGTGCTGCGCGCCCTTCTACGTGTGATTTGGCGCGCTGCGCCA 760
Db 1979 GGGTCGCGCACCCAGTACAGTTCACTTTCAGGAGCGCGACCCGAGCTGTCTCGCGC 2038
Qy 761 GCGTTGGCCACGCCCTGCGAGGAGCGGTGAGCGCGCGCCCATGTTCGCGCCACCT 820
Db 2039 AGTGGGTGCGCAAGCTGTGCGCGCTGCGAGGTTGCGCAGCTGCGCGAGCTCGGCCA 2098
Qy 821 TCCTGTACACGCACTGTGGGAGGACCCCGAGCGGATATGAGAGTGTGATGAGATCAACC 880
Db 2099 GCGACTGGCAGGACAGGGCTTGCAGGCCCTTCTGAACTATCGACCGCG-----ACACCG 2152
Qy 881 CCAAGGACACGGTGTGACCTTACGTAGCGCGCGCTGCAATGCCCTGAACTGCTGTGTGC 940
Db 2153 CTTGCGGCTTCCGGCTGAAGCTCTCGACATCGACAGCGTGTCTTACAAGCGCTTCGGCC 2212
Qy 941 AGGGGCGCGCAGGTGTGCTGGTGGACTGCAACCCCGCGAGTCCGCGCTTCTGAGC 1000
Db 2213 AGCGGTGTATCTCGACATCTTCAACGAGCCACCCAGTACCCTGCTGTGAGGTGG 2272
Qy 1001 TGAAGAGGTGGCCATTCAGCAGCTGGAGTTTGAAGACGTGTGTGAGCTGTTCGGCGAGG 1060
Db 2273 CGCGCAGTTCCAGCTCGGCCCGCGCCCTTGGAGCAGCTCTAGTCCGCTCCAGCGACG 2332
Qy 1061 GCGTGACCCCGCGATTTGAGAGCTGTACGAGAAGAGCTGGCGCCCTTCTGTGCGAAA 1120
Db 2333 GCACCCAGGTGCGCTGTGAGCTTGGCGAAGGTGGAGAGCGGCATACCTTGTGCGCA 2392
Qy 1121 CGAGCCACACTTCTGTGTCAGCGCTCTGTACTTCCAGCAGCGCTGTACTACAGG 1180
Db 2393 TCAACCATATCGCCCAAGTTC---CCCTCGGCGACCTGTCTTCAACTGGCCAGGGTT 2449
Qy 1181 GCGCATGCGCAAGCTGTGTGCTGCTGAGTGCCTTGGCGCTGTGCTGCGACTGGGCA 1240
Db 2450 ACTCCCTGGGAGGCGGTGAGGCGATCCGTGCGCTCGAGGCCAGCTTGGAGCTGCGCG 2509
Qy 1241 AGACCGTCAAGCGCTTCGCAACCGGCCCAATGAGAGGAGCAGCGCGCTGTGTGGGACA 1300
Db 2510 TGAGCATGACGCGCAGCTTCCGCGCGCGCGCTTGGCTTTCGAGGCGCTCGTGTGCAACA 2569
Qy 1301 GCAACATGCTCATCC---ACTTGTGAGAACGGGCCCAAGCGCTGTGTGCTGTGCTG 1357
Db 2570 CGTGTGCTGTATCTCGCTTCCGCTGTGTGAGCATGTATCATGCTTGGGCACTCTTACG 2629
Qy 1358 TCAAGTTCTGTGAGCTGTGTCTTCAACAGGCGCTGTGTGCTTGGCGCGCGCGCTGC 1417
Db 2630 AGAGTTTCATCCATCCCGTGAACCATCTCTCGACCTGCTCGCGCGGGGTGCGCGCG 2689
Qy 1418 CGGCAAGCAGTACGCGCTGTATCAAGCGGAGCGGCATCCCATTTGAGAACTACATCGGCG 1477
Db 2690 TGCTGGCGCTGATGCTGCGCGGCGAGGAGATCGGCATCGTGGCGCATCATCGGCATC 2749
Qy 1478 GCACCATGAGCGGTGCGGAGAACTCGCACGTGCGCAGCAGAACTACTTCTACTACA 1537
Db 2750 TGCTGATCGGCATCTGTCAAGAAAGACCGCATCATGATGATGATGCTTCCGCTTCAGCGCG 2809

QY 1538 ACTGCTCAGCGCAAG 1554
Db 2810 AGCGCAACGAGGCAAG 2826

RESULT 3

US-09-252-991A-14823/c

; Sequence 14823, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; PRIOR FILING DATE: 1999-02-18

; PRIOR FILING DATE: 1998-02-18

; PRIOR FILING DATE: 1998-02-18

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 14823

; LENGTH: 3411

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-14823

Query Match 3.6%; Score 69.4; DB 3; Length 3411;
Best Local Similarity 43.8%; Pred. No. 0.00015;
Matches 454; Conservative 0; Mismatches 571; Indels 12; Gaps 3;

QY 521 TGGCGGTTCGCGACTTCTACGTGAGCGGCAAGTACGACCTGCCCTCGCGCAGATGCCCT 580
Db 1349 TCGGCGTCGACGCGCAGCAACCCGACCTCAACACCGCGCGCTGCTGATCAACTCAAGC 1290
QY 581 GGTGCGCGCTTTCTTCTGCGGATCGATCTTCGACATCGACAATGACATCGGCGCG 640
Db 1289 CGCACAGCAGCGCGAGCTCACCGCCAGCGAAGTATCCAGCGCTGCAGCGCGAATCG 1230
QY 641 AGCGCGCGCTTACCTTGAGCAGAGCTGGAGCGCGTGGAGCGGAGCAGACACCCAGGTT 700
Db 1229 ACACCTGCCGGATCAAGCTGTACATGACGCGGTGAGGACCTGACCATCGAGACC 1170
QY 701 CGATCCCTACGTGCGGTGCTGCGCGCCCTCTACTACGTGTGATGGCGCGCTGCCCA 760
Db 1169 GGGTCGCGCGCACCCAGTACAGTTTACCTTGCAGGAGCGCGACCGAGCTGCTGCGG 1110
QY 761 CGGTGCGCACGCGCTGCAGAGAGCGGTGAGCGCGCGCCCATGTTCCCGCCCACT 820
Db 1109 AGTGGGTGCGCAAGCTGTGCGCGGCTGCAGGAGTTGCCGCGAGCTCGCGAGCTGCCA 1050
QY 821 TCCTGTACACGCAAGTGTGGAGGAGCCCGCAGCGGATATGGAGGTGATGGAGATCAACC 880
Db 1049 GCAGCTGGAGGAGCAAGGGCTTGAGGCTTACTGAACTGACCGG-----ACACCG 996
QY 881 CCAAGGACAGGTGCTGACCTGACTAGCGCGGCTGCAATGCCCTGAACTGCTGCTGTC 940
Db 995 CTTGCGCGCTCGCGTGAAGCTCTCCGACATCGACAGCGGTCTACAGCGCTTCGCGC 936
QY 941 AGGGGCGCGCAGGTGTGCTGCTGAGTGTGACATCGACCGCGAGTGGCGGCTTCTGAGC 1000
Db 935 AGCGGCTGATCTCGACCATTTTCAACCGAGCGCCAGTACCGCGGTGCTGAGGAGTG 876
QY 1001 TGAAGAAGTGGCCATTTACAGCAGCTGAGTTTGAAGGAGCTGTGCGAGCTGTTGCGGAGG 1060
Db 875 CGCGCAGTTTCAAGCTCGCGCGCGAGCGCTTGAGGAGCTTACGTGCGGTTCAGCGAGC 816
QY 1061 GCGTGACCGCGCATTTGAGGAGCTGTACGAGAGAGAGTGGCGCGCTTCTGCTGCAAA 1120
Db 815 GCACCGAGGTGCGCTGCTGAGCTTGGCGAGGTGGAGGAGCGCATACCTGCTGCGCA 756
QY 1121 CGAGCAGCAACTTCTGCTGCAAGCGCTCTGCTGATTTCCAGCAGCGCGCTGTATACAGG 1180

Db 755 TCAACCATATCGCCCAAGTTTC---CCCTCGCGGAGCCCTGTCTGTTCAACCTGCGCCCAAGGTT 699
QY 1181 GCGGCATGGGCAAGCTGTGCTGGGTGCTGAGTGCCTGCGCGCTGCTGCTGCGACTGGGCA 1240
Db 698 ACTCCCTGGCGGAGCGGTCGAGCGGATCGTGGGTGAGGCCAGCTTGGAGCTGCGCG 639
QY 1241 AGACCGTCAAGCGCTCGCCCAACGCGCCCAATATGAGAGAGCAGCGCGCTGCTGAGGACA 1300
Db 638 TGAGCATGCAAGGCGAGCTTCCGCGCGCGCGCTGCGCTTTCGAGGCGCTGCTGTCGAACA 579
QY 1301 GCAACATGCTCATCC---ACTTCTGTAAGAAACGGGCCCAAGCCGCTGCTGCTGCTGCTG 1357
Db 578 CGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 519
QY 1358 TCAAGTTTCTGAGCTGCTGCTGCTTCAACAAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 1417
Db 518 AGAGCTTCAATCCATCGGTCGATCATCTCTCGACCTGCTGCTGCTGCTGCTGCTGCTGCTG 459
QY 1418 CGGCGAAGCAGTACGCGCTGATCAAGCGGAGCGGATGCCATTCCTTGAAGTATGAGAACTACATCGCG 1477
Db 458 TGCTGCGCTGATGCTGCGCGGCGGAGATCGGATCGTGGCGATCATCGGATCATCC 399
QY 1478 GCACCATGAGCGGCTGCGGAGAACTCGCACCTGCGCAGCAGCACTTCTTACTACA 1537
Db 398 TGCTGATCGGATCGTCAAGAAACGCGGATCATGATGATGATGATGATGATGATGATGATG 339
QY 1538 ACTGCTCACCGGCAAG 1554
Db 338 AGCGCAACGAGGCAAG 322

RESULT 4

US-09-010-928B-1

; Sequence 1, Application US/09010928B

; Patent No. 5994099

; GENERAL INFORMATION:

; APPLICANT: Lewis, Randolph V

; APPLICANT: Hayashi, Cheryl Y

; TITLE OF INVENTION: EXTREMELY ELASTIC SPIDER SILK PROTEIN AND DNA

; TITLE OF INVENTION: CODING THEREFOR

; NUMBER OF SEQUENCES: 29

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH

; STREET: 8110 GATEHOUSE RD. SUITE 500E

; CITY: FALLS CHURCH

; STATE: VIRGINIA

; COUNTRY: UNITED STATES OF AMERICA

; ZIP: 22042

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; FILING DATE: 22-JAN-1998

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Murphy Jr., Gerald M

; REGISTRATION NUMBER: 28977

; REFERENCE/DOCKET NUMBER: 1447-109P

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2830 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: not relevant

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; HYPOTHETICAL: NO

; FEATURE:

; NAME/KEY:

; LOCATION: 1..2830

; OTHER INFORMATION: /note= "Flagelliform DNA sequence"

OTHER INFORMATION: taken from the 5' region. The putative start codon is at
OTHER INFORMATION: position 219"

FEATURE:
NAME/KEY: CDS
LOCATION: 219..2830
US-09-010-928B-1

Query Match 3.5%; Score 68.2; DB 2; Length 2830;
Best Local Similarity 46.4%; Pred. No. 0.00025;
Matches 290; Conservative 0; Mismatches 333; Indels 2; Gaps 2;
QY 772 GCCTTCACGAGAGCGCTGGAGCGCGCCCATGTTCCCGCCACCTTCCTGTACACG 831
DB 1033 GACCTGTGGAGCGCGACACAGAGGTTATGACCTGTGTGGAGCGCGACAGGAGTTACG 1092
QY 832 CAGTGTGGAGAGACCCCGAGCGGATATGAGAGTGATGAGATCAACCCCAAGAGACAG 891
DB 1093 GACCTGTGGAGCTGGACACAGGAGTTACGACCTGTGTGAGCTGGGCTTGGAGGTTACG 1152
QY 892 GTCTGACCTGACTAGCGCGGCTGCAATGCCCTGAACTGCTGTGTGACAGGGCGCGGC 951
DB 1153 GACCTGTGGAGCTGGACCTGGAGGTTACGACCTGTGTGAGCTGGACCTGGAGGTTACG 1212
QY 952 CAGTGTGTCTGGTGAAGTCAACCCCGCAGTTCGCGCTTCTGGAGCTGAAGAGGTTG 1011
DB 1213 GACCTGTGGAACTGGACCTGTGTGATACGACCTGTGTGAACTGGACCTGGAGGTTG 1272
QY 1012 GCATTCAGCAGCTGGAGTTTGGAGAGCTGTGGAGCTGTTCGGCGAGGCGGTGCACCG 1071
DB 1273 GACCTGGAGAGCTGGACACAGGAGATATGACCTGTGTGTGTGGACCTGTGTGTGCTG 1332
QY 1072 CGCATTTAGAGCTGTACGAGAGAGAGCTGCGCCCTTCCTG-TCCGAACACGCCACAA 1130
DB 1333 GACCTGTGTGTGTGGACCTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 1392
QY 1131 CTTCTGTGTCAAGCGCTCTG 1190
DB 1393 GACCTGTGTGTATACGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 1452
QY 1191 CAAGCTGT 1250
DB 1453 GACTTGTGTGT-GCTGT 1511
QY 1251 GCGCTCGCCNACGCGCCCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1310
DB 1512 GGAACAGT 1571
QY 1311 CATCCACTTGTGAAGAACCGCGCCCAAGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTG 1370
DB 1572 GGACCCGCTG 1631
QY 1371 CTTGTGTGTCTTCAACAGGCGGTG 1395
DB 1632 GGTG 1656

RESULT 5

US-09-902-540-6564
; Sequence 6564, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 6564

LENGTH: 1691
TYPE: DNA
ORGANISM: Myxococcus xanthus
US-09-902-540-6564

Query Match 3.2%; Score 63; DB 3; Length 1691;
Best Local Similarity 44.9%; Pred. No. 0.0024;
Matches 323; Conservative 0; Mismatches 390; Indels 6; Gaps 2;

QY 659 AGCAGAAGCTGGAGCGCGTGTGGAGCAGACACACCCAGGTTTCATCCCTACGTCGCGT 718
DB 260 AGGTGCGCTGGCGCACCGTGGCGCGAAGCTGCCAACCGCATGAGCATCAAGAGCGCG 319
QY 719 GGTTCGCGCGCCCTACTAGTGTGATTTGGCCCGCTGCCACGCTTGGCCAGCCCTGTC 778
DB 320 CGCTGCGCGCGTGGACAGCTTCGGCATGCTCTGCTCTCTCA--AGGAGCTGGGCTGA 376
QY 779 ACAGAGAGCGGTGGAGCGCGCCCATGTTCCCGCCACCTTCTCTGTATACGAGTCGT 838
DB 377 GCGAGAGTCCAGCGGCTCTCATCTTCCGCGGACACGCGCTGGGACGCGCCATCG 436
QY 839 GCGAGACCCCGAGCGGATATGAGGTGATGAGATCAACCCCAAGGACACGTCGTGA 898
DB 437 CGAGAGCGGTGGGCTGGATGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 496
QY 899 CCCTGACTAGCGCGCTGCAATGCCCTGAACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 958
DB 497 ACCTGCTACGACCTGTGGCGGTGGCGGAGGTGAGCGTGTGTGTGTGTGTGTGTGTGTG 556
QY 959 TGTTCGTGTGACTCAACCCCGCGAGTTCGCGCTTCTGTGTGTGTGTGTGTGTGTGTGTGTG 1018
DB 557 AGTGTCCCGACGCGGAGTTCGCGGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 613
QY 1019 AGCAGCTG 1078
DB 614 TGCCTG 673
QY 1079 AGGAGCTG 1138
DB 674 AGATTGCGCGCTTCCGAGTG 733
QY 1139 CCAAGCGCTTG 1198
DB 734 TCAACAGCTG 793
QY 1199 GCTGGGTG 1258
DB 794 CTTGCACTG 853
QY 1259 CCAAGCGCCCAATG 1318
DB 854 AGAAGCTCAAGCTG 913
QY 1319 TCGTGAAGAACGCGGCCCAAGCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 1377
DB 914 CGGACAGGACCGCGCGCGCCATCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 972

RESULT 6

US-09-902-540-8514
; Sequence 8514, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 8514
; SEQ ID NO 8514

```

; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 8514
; LENGTH: 2034
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-8514

Query Match      3.2%; Score 63; DB 3; Length 2034;
Best Local Similarity 49.5%; Pred. No. 0.0025;
Matches 162; Conservative 0; Mismatches 165; Indels 0; Gaps 0;

QY 790 GTGGAGCGCGCGCCATGTTCCCGCCACCTTCTGTACAGCGAGTCTGTGGAGGACCC 849
D 790 GTGGAGCGCGCGCGCCATGTTCCCGCCACCTTCTGTACAGCGAGTCTGTGGAGGACCC 849
DB 664 GAGGAGCGCGGCTTCTGTCTGTGGAGCGAGTCTGTGGAGCGCGG 723
QY 850 GAGCGGATATGAGGAGTGTGAGATCAACCCCAAGGACACGCTGTGACCTCTGACTAGC 909
D 724 GCCTGTACGAGGCGCTGCTGTGAGCGCGGCAAGCTGCAGCTGTGGCGCTGGAGCGGCG 783
QY 910 GCGCGCTGCAATGCTGTGAACCTGTGTGGAGCGCGCGGCGAGGCTGTGCTGTGGAG 969
D 784 AAGGAGCGCGGAGGAGCGGCTTCTTCTGTGTGAGCGAGCGCGGCTGTGGCGGAGCAC 843
QY 970 TGCAACCCCGCGAGTCTGTGAGCTGAAGAGTGGCCATTGAGAGCTGGAG 1029
D 844 CTGGAGCGAGTGTGAGTGTGAGCGCTGTGGAGCACCTGTGGCGAGTGGCGCGGAAACGTGGCC 903
QY 1030 TTTGAGGACGTGTGGAGCTGTTCGGCGAGGCGGTGCACCGCGCATTTGAGGAGCTGTAC 1089
D 904 ACCTACTGCTGTGTGAGTGTGAGCGAGTGTGAGCGAGCGGCGTGTGGCGCGGCTGTGGCGGCTGC 963
QY 1090 GAGAAGAGTGTGGCGGCTTCTGTGTG 1116
D 964 CACGCGGAGCGAGCGGAGCTGTG 990

RESULT 7
US-09-902-540-489
; Sequence 489, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; PRIOR FILING DATE: 2001-07-10
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 489
; LENGTH: 2730
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-489

Query Match      3.2%; Score 63; DB 3; Length 2730;
Best Local Similarity 44.9%; Pred. No. 0.0026;
Matches 323; Conservative 0; Mismatches 390; Indels 6; Gaps 2;

QY 659 AGCAGAGCTGGAGCGCGTGTGGAGCAGAACCCAGAGGTTCTGATCCCTCTAGTGGCGGT 718
D 1299 AGTGCGCTGTGGCAGCGTGTGGCGGAGAGTGTGCGCAAGGATGAGATCAACAGCGCGG 1358
QY 719 GGCTGCGCGCGCTTACTAGTGTGATTTGGCGCGCTGTGGCGAGCTGTGGCGCGCTGTG 778
D 1359 CGCTGCGCGCGTGGAGCAGCTTCTGGCATGTCTGTCTGTCTCA--AGGAGCTGGCGCTGA 1415
QY 779 ACAGAGCGCGTGGAGCGCGCGCATGTTCCGGCGCACCTTCTGTGTACAGCGAGTGT 838
D 1416 GCAGGAGTCTCAGCGCGCTCTCATCTTCCGCGGAGACAGCGCGTGGGCGACGCCCATCG 1475

; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 8514
; LENGTH: 2034
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-8514

Query Match      3.2%; Score 63; DB 3; Length 2034;
Best Local Similarity 49.5%; Pred. No. 0.0025;
Matches 162; Conservative 0; Mismatches 165; Indels 0; Gaps 0;

QY 839 GGGAGGACCCCGAGCCGGATATGGAGGTGATGAGATCAACCCCAAGGACACGGTGTCTGA 898
D 1476 CGGAGCGCGTGGCGCTGGATGGCGTGTGAGGTGAACGTCACGCCGAACCGCCCGG 1535
QY 899 CCTGACTAGCGCGCGCTGCAATGCCCTGACCTGTGTGTGTCAGGGGCGCGGCGAGTGG 958
D 1536 ACAGCGCTCAGCCACCTTGGCGGTGGCGCGAGGTGAGCGTGGTGAACGGGCTCCGCGTTGA 1595
QY 959 TGTGCGTGGACTCAACCCCGCGCAGTCCGCGCTTCTGTGAGCTGAAGAAGTGGCCCATTC 1018
D 1596 AGTGCGCGCGCGCGCGAGCGCGGAGTCTGGGACACGCCCGCGCGGAGCAGGTG--AAGG 1652
QY 1019 AGCAGCTGAGTGTGAGGAGCTGTGGAGCTGTTCGGCGAGGCGGTGCAACCGCGCATTTG 1078
D 1653 TGGCGCTGGAGGCGCGCGACCGCTGCGCGCTACGTGGCGCGCTGTGGTGGAGAACGTGA 1712
QY 1079 AGGAGCTGTACGAGAGAGAGCTGGCGGCTTCTGTGCGAAACCGAGCCACACTTCTGCT 1138
D 1713 AGATTGGCGCGCTTCCGCGAGTGTGAGGACCGGCTGAAGGCGCGCGGCTGGCGGCCA 1772
QY 1139 CCAAGCGCTCTGTGTACTTCCAGCAGCGCTGTACTACAGGCGCGCATGGGCAAGCTGT 1198
D 1773 TCAACACGTGGTGGATGTACCAACTACGTGAACCTGGAGTACGGGACCGCGCTGACG 1832
QY 1199 GTTGGTGTGCTAGTGTGCTGGCGCTGTGGGACTGGGCAAGACGTTCAAGCGCTCG 1258
D 1833 CCTTACCTGGAGAGCTGGCGGCTGAGGAGTGTGCTGCTGCGTACGCGGACCGCGCGG 1892
QY 1259 CCAACCGCGCCACAAATGGAGGAGCAGCGGCTGTGGGACAGCAACATCTATCCACT 1318
D 1893 AGAAGCTCAAGACGCTGGACGCGCAAGGACCGCGCTCTGGAGCTGGATGACCTGTCTATCG 1952
QY 1319 TCGTGAAGAACGCGGCCCAAGCGCTGTGTGGCTGTTCGTCAGATTGCTGAGCGCTGTG 1377
D 1953 CGGACAGGACCGCGCGCGCATCTGTGGCGCTGTGGCGCGCGGCGGACAGCGAGGTG 2011

RESULT 8
US-09-902-540-897
; Sequence 897, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; PRIOR FILING DATE: 2001-07-10
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 897
; LENGTH: 6855
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-897

Query Match      3.2%; Score 63; DB 3; Length 6855;
Best Local Similarity 49.5%; Pred. No. 0.003;
Matches 162; Conservative 0; Mismatches 165; Indels 0; Gaps 0;

QY 790 GTGGAGCGCGCGCCATGTTCCCGCCACCTTCTGTACAGCGAGTCTGTGGAGGACCC 849
D 665 GAGGAGCGCGGCTTCTGTCTGTGGCGAGCAGTGTGGCGCGCGCTGTGGAGCGCGG 724
QY 850 GAGCGCGATATGAGGAGTGTGAGATCAACCCCAAGGACACGCTGTGACCTCTGACTAGC 909
D 725 GCCTGTACGAGCGCTGTGAGCGCGCGGCGGCAAGCTGACGCTGTGGCGTGGAGGCGGCG 784
QY 910 GCGCGCTGCAATGCTGTGAACCTGTGTGTCAGGGGCGCGGCGGAGTGTGTCGGTGGAC 969
```

Db 785 AAGAGCCGAGGAGCGCTCTTCTCTCTGAGCGAGCGCGCTGCGGAGCAC 844
Qy 970 TGAACCCCGGAGTGGCGCTTCTGAGACTGAGAGGTGGCCATTGAGAGCTGGAG 1029
Db 845 CTGGACACGTGAGTGGAGCTGGAGCACCTGGCGAGTGGCGCGGAGAACTGGCC 904
Qy 1030 TTTGAGGACGTGGAGCTGTTGGCGAGGGGTGCACCGCGCATTTGAGGAGCTGTAC 1089
Db 905 ACGTACTGCTGTGGAGTTGATGGAGCGGAGCGGTGCGCGCGCTGGCGCGTCC 964
Qy 1090 GAGAAAGAGCTGGCGCTTCTCTGTCTG 1116
Db 965 CACCGGAGCCAGAGCGGAGCTGTG 991

RESULT 9

US-09-881-165-4
; Sequence 4, Application US/09881165
; Patent No. 6632930
; GENERAL INFORMATION:
; APPLICANT: HOOD, ELIZABETH
; APPLICANT: HOWARD, JOHN
; APPLICANT: BAILLEY, MICHELE
; APPLICANT: GASTEL, FRANS VAN
; APPLICANT: WANG, HUAMING
; APPLICANT: WARD, MICHAEL
; APPLICANT: WOODARD, SUSAN
; TITLE OF INVENTION: METHOD OF INCREASING RECOVERY OF HETEROLOGOUS ACTIVE
; FILE REFERENCE: 10032R
; CURRENT APPLICATION NUMBER: US/09/881,165
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: 60/211,732
; PRIOR FILING DATE: 2000-06-15
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 1082
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic DNA encoding
; OTHER INFORMATION: Organophosphate Hydrolase
US-09-881-165-4

Query Match 3.2%; Score 62.4; DB 3; Length 1082;
Best Local Similarity 44.6%; Pred. No. 0.003;
Matches 246; Conservative 0; Mismatches 306; Indels 0; Gaps 0;
Qy 1371 CTGGTGCTCTTCAAGCGCGTGTGTGTTGGCGGGCGGTGCGGGGCAAGTGA 1430
Db 515 CCAGGAGCTGTGCTCAAGCGCGCGCGCTCTCCCTCGCCACCGCGGTGCGGTGAC 574
Qy 1431 CGCGCTGATCAAGCGGAGCGCATCCCATTTGACAATACATCGCGCGCATGAGCGG 1490
Db 575 CACCACACCGCGCTCCAGCGGAGCGGAGAGCGGCGCCCATCTTCGAGTCCGA 634
Qy 1491 CGTGGCGAGAACTCGCACGTGCGCAAGCAAACTACTTCTATACAACTGCTTCAACGG 1550
Db 635 GGGCTCTCCCGTCCCGGTGTGATCGCGCACTCCGAGACACCGACGACCTCTCTTA 694
Qy 1551 CAAGTTCGTGGGAGCAAACTGCCCCCACTTACTCGGAGGCGGCTTGGCCACCTCAA 1610
Db 695 CCTCACGCGCTCGCGCGCGGCTACTCTATCGGCTCGACACATCCCGCACTCCGC 754
Qy 1611 GAGTGGGTGTGAGCAACCTGACCGTCTCCACCAACTTCTTCAATGAGGAGCTCAAAGC 1670
Db 755 CATCGGCTCAGAGCAACGGCTCGGTGCGGCTCTCTGGCATCGCTCTGCGAGAC 814
Qy 1671 GCGCACCTTACCAAGGTGATTTGATGGACACGTGAGTGGGTGATGCGCGGTGC 1730
Db 815 CCGCGCTCTCTATCAAGGCGCTCATCGACGAGGCTTACATGAAGCAGATCTCTGTC 874

Qy 1731 CAACGAGCTGGCGAGTGTGCTGGCCAAAGCAGAGTGTGGCCGCGGCGCATCGTCTATCTGGCG 1790
Db 875 CAACGAGTGGCTTTCGGGCTTCTCTCTTACGTGACCAACATCATGAGCTGATGGACCG 934
Qy 1791 CTCGGCTCTCTCAGCCGCGCTTACCGCGAGCTGATTCAGAGAGCGGGCTTTCGACGTGGCG 1850
Db 935 CGTGAACCCGAGCGCATGGCTTCTATCCGCTCCGCTGATCCCGTTCCTCCGCGAGAA 994
Qy 1851 CTGCATCCCGCGCGCACTCAGGGCTACATGGACCGGCTCAACATGTACAGCTCTCTTA 1910
Db 995 GGGCGTGGCGAGGAGACCTCGCGGCTACCGTGAACCAACCGCGCGCTTCTCTCTC 1054
Qy 1911 CATGGCCCGCG 1922
Db 1055 CCCGACCTCCG 1066

RESULT 10

US-09-902-540-2993
; Sequence 2993, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 2993
; LENGTH: 1167
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-2993

Query Match 3.2%; Score 62.4; DB 3; Length 1167;
Best Local Similarity 43.7%; Pred. No. 0.003;
Matches 340; Conservative 0; Mismatches 426; Indels 12; Gaps 1;
Qy 762 CGTTGGCCAGCCCTGCACGAGGCGCGTGGAGCGCGCCCATGTTCCCGCCACCTT 821
Db 351 CTTGAGACTACGCGCGCGCAGAGTTTCGTGAGTGAGACCCGCTGCTATGCGCTACT 410
Qy 822 CTTGTACACGACGTCTGTTGGAGGACCCCGAGCGGATATGGAGTGTATGGAGATCAACCC 881
Db 411 GCGGCGCAGCATGCGCGCGGAGCAGCGGAGCTGTGCGCAGCCTGAAGATCATCCAGGA 470
Qy 882 CAAGGACACGGTGTGACCTGACTAGCGCGGTGCAATGCGCTGAACTGCTGTGGTGCA 941
Db 471 CGACATCACGTACATACCTGCGCGCGCTGAAGGCTGACCTCAACGTGCGCGCT 530
Qy 942 GGGGGCGCGCAGTGTGCTGCTGAGTGTGCTGAGTGTGCTGAGTGTGCTGAGTGTGCTGAGT 1001
Db 531 GGGCTCGGGCTGGAGGGCGCGCTGCGCTTCGACTTCAGCCACGTCAGCAAGGTGGAGCC 590
Qy 1002 GAGAGAGGTGGCTATTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGT 1049
Db 591 GAGGGCTTCCGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGT 650
Qy 1050 GTTGGCGAGGGGTGACCGCGGCTATGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGT 1109
Db 651 GTTGTGCGGTGCGCGCGCTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 710
Qy 1110 CTTGTGCAAAACAGCCACAACTTCTGCTCCAGCGCTCTGCTACTTCCAGACGCGCT 1169
Db 711 GGTGCGGTGCGCCACGCTGTGGCTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 770
Qy 1170 GTACTACAGGGCGCATGGGCAAGCTGTGCTGGTGTGCTGCAAGTGTGCAAGTGTGCAAGTGTGCAAGT 1229

Db 771 CCAGCGCTCAAGCCGCGGACTACTTGGCAGGCTCGGGCGAGCGGCTCCGCGGTGA 830
Qy 1230 GGAGCTGGGAAGACCGTCAAGCGCTTCGCAACGCGCCCAACAATGAGAGAGCAGCGCG 1289
Db 831 GGTGCGGTGCCCCATCTGCGCGCGCAACGCGCGTCCCCACATGCCCCAGCTCCAGGG 890
Qy 1290 TCTGTGGGACGACCAATGCTCATCTCAGCTTGTGTAAGAACGGGCGCCAGCGCTGTGTG 1349
Db 891 GCTGCTGGCGCGGTGAGCTGAGCTGAGCGCGCGCTGGAGGAGCTTGGAGCGCTGGAGAC 950
Qy 1350 GCTGTTCTGCTCAAGTTCTGAGGCTGTGTCTTCAACAAGCGCGTGTGTGTTTCGCGG 1409
Db 951 TCGGCGCTCAGTCAGTACCTTTCGGCGGCAAGATCTGACCCGCGCGCGCTCAGGG 1010
Qy 1410 CGGCGTTCGGGCGCAAGCAGTACGCGCTGATCAAGCGCGGAGCGGATCCCAATTGAGAACTA 1469
Db 1011 CGCGTCCAGCGACATCTCCAACTCGCTGGCGAGCACCAAGCTGAACATCATCCGCGGCT 1070
Qy 1470 CATCGCGCGCACCATGACGCGGTGGCGGAGAACTCGCAGTGGCGGAGCAGAACTAC 1527
Db 1071 GGGACAGGGCGGATGGCGAGGTCTTCTCGCAAGCAGGTGGCGGTGAAGGGCTTC 1128

RESULT 11

US-09-902-540-362
; Sequence 362, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 362
; LENGTH: 1250
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-362

Query Match 3.2%; Score 62.4; DB 3; Length 1250;
Best Local Similarity 43.7%; Pred. No. 0.003;
Matches 340; Conservative 0; Mismatches 426; Indels 12; Gaps 1;
Qy 762 CGTTGGCCAGCCCTGTCACGAGGAGCGCGTGGAGCGCGCCCATGTTCCCGCCACCTT 821
Db 434 CTTGAGTACGCGGGCGGCGAGCAGTTCGTGTCAGGTGGACCGCGTGTATGCGTACCT 493
Qy 822 CTTGTACCGCAGTGTGTGGAGACCCCGAGCGCGATATGAGGTGATGAGATCAACCC 891
Db 494 GCGCGCGCAGCATGTCGCGCGGAGCAGCGGAGCTGTGCGACACCTGAAGATCATCCAGGA 553
Qy 882 CAAGGACACGGTGTGACCTGTAGCTAGCGCGGCTGCAATGCTGCACTGTGTGTGCA 941
Db 554 CGACATCACGTATACATCACCTGTCGCGCGCTGAAAGGGTGACTCAACGTGCGCGCT 613
Qy 942 GGGGCGCGCGCAGTGTGTGCTGTGACTGCAACCCCGCGCGCTGCGCGCTTCTGGAGCT 1001
Db 614 GCGTCTGGGCTTGAGGGCGCGCTGCGCTTTCGCTTTCAGCCAGTCAAGAGTGGAGCC 673
Qy 1002 GAAGAAGTGGCCATTACAGAGCTGGAGTTTGAAGACGTTGGCA-----GCT 1049
Db 674 GGAGGCCCTCCGAAGCTGAGCAGGTGTGAGAGCGGCGCGCAGCGCGCGCACGTGGT 733
Qy 1050 GTTCGGGAGGGCGTGCACCCGCGCATTTGAGGAGCTGTACGAGAAAGCTTGGCGCCCTT 1109
Db 734 GTTGTGCGGGTGCAGCGCGCGCTGTTGGCGGTGCTGGCGCGCTCGGCCAAGGTGTTGCC 793

Qy 1110 CTTGTGCGAAACACAGCCACAACTTCTGTGTTCCAGCGCGCTTGTGTAATTCAGCAGCGCCT 1169
Db 794 GGTGCGGCTGGCCACAGCTGTGGCTGCGCTGTGACTGCTCCGCAACTGCGGGCAGGTGAGCCA 853
Qy 1170 GTACTACAGGGCGGCGATGGCAAGCTGTGCTGGGTGCTGCACTGTGCTGCGCGTGTGCT 1229
Db 854 CAGCGCTTCCAGCGCGGACTTACTTGGCAGGCTGCGGGCGAGGCGTTCGCGGTGGA 913
Qy 1230 GGGACTGGGCAAGACCGTCAAGCGCTCGCAACGCGCCCAATGAGAGAGCAGCGCGC 1289
Db 914 GGTGCGCTGCCCATCTTGGCGGCAACGCGCGCTGCCCATATGCCAGCTCCAGGG 973
Qy 1290 TCTGTGGGACAGCAATGCTCATCTTCTGTGAAGAACGGGCGCCAGCGCTGTGTG 1349
Db 974 GCTGCTGGCGCGGTGAGCTGACGACCGCGCGCTGGAGGACTTGGAGCGCTGGAGAC 1033
Qy 1350 GCTGTTCTGCTCAAGTTCTGTGAGCTGTGCTTTCACAGGCGCTGTGTGTTTCGCGG 1409
Db 1034 TCGCGCGCTCAGTCACTTCTTTCGCGCGCAGCAACATGACCCGCGCGCGCTCAGGG 1093
Qy 1410 CGGCGTGGCGGCGAAGCAGTACGCGCTGATCAAGGCGGAGCGGATCCCAATTGAGAACTA 1469
Db 1094 CGGCTCCAGGACATCTCCAACTCGCTGGGCGACCAAGCTGAACATCATCCGCGGCT 1153
Qy 1470 CATCGCGCGCACCATGAGCAGCGGTGGCGGAGAACTGCGAGTGGCGGAGCAGAACTAC 1527
Db 1154 GGGACAGGGCGGCGATGGCGAGGTCTTCTCGCAAGCAGGTGGCGGTGAAGGGCTTC 1211

RESULT 12

US-09-902-540-7318
; Sequence 7318, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 7318
; LENGTH: 3579
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-7318

Query Match 3.2%; Score 62.2; DB 3; Length 3579;
Best Local Similarity 42.3%; Pred. No. 0.003;
Matches 474; Conservative 0; Mismatches 638; Indels 9; Gaps 2;
Qy 821 TCTGTACACGCGAGTCTGTGGAGACCCCGAGCGCGATATGAGGTGATGAGATCAACC 880
Db 1517 TCCAGAGCCCAAGGTGTGTGGAGCCCAAGGACGACAAAGGCGGAGCGGCGGCGAGC 1576
Qy 881 CCAAGGACACGTTGTGACCTTGTGACTAGCGGCGGTGCAATGCTGTAACCTGCTGTGC 940
Db 1577 CCGGCAAGTGGTGTGAGAACCGCGCTACAAACGCGGTGTGGCTTACGACGAGTGTCA 1636
Qy 941 AGGGGCGCGGCGAGTGTGTGCTGAGTGTGCAACCCCGCGAGTGGCGCTTCTGGAGC 1000
Db 1637 AGCGCGCGGAGCGCGGCGGCGGAGGCGGAGCGGCGGTGGGCTCGGACATCACCAGA 1696
Qy 1001 TGAAGAAGTGGCCATTACAGCAGTGTGAGTGTGAGGACGTTGTGGAGCTGTTCGGCGAGG 1060
Db 1697 AGGCCACCATCCCGACGCTGAAGAGCGCTGTCTCAGCGGTGTGAGCGCTACTCAAGT 1756
Qy 1061 GCGTGACCCGCGCATTTGAGGAGCTGTACGAGAAAGCTGGCGCGCTTCTCTGTGCAAA 1120

```
Db 1757 ACCTGCGCAAGAGGTGAGAGCGGTGGAGATCGCTTCAAGGGCGCCAAATCTACTTACC 1816
Qy 1121 CCAGCCCAACTTCTGTCTCAAGCGCTCTGTACTTCCAGACGCGCTGTACTTACCAGG 1180
Db 1817 GCACAAACCACTTCGACGAGCGGTGTCTGCGCTTCAGCGAGATCGCGCTCGGTACCCCG 1876
Qy 1181 CGGCGATGGGCAAGCTGTCTGGGTGTGTCAGTGCCTTGGCGGTGTGTCGGAGCTGGCA 1240
Db 1877 AGTACAAGTTCGAGAACGGCGGCGCGGAGATCTCGCCAACTCATCTCGACT 1936
Qy 1241 AGACCGGTCAAGCGCTCGCCAAAGCGCCCAATATGGAGGAGACGCGCTGTGTGGACA 1300
Db 1937 CGTACCACTCTGTGACGACTACGGAAGGTGAACGAGTGGCGCGCGCTTCTACGCCA 1996
Qy 1301 GCACATGC-----TCATCCACTTCGTGAAGAACGGCGCCCAAGCGCTGTGTGGTGT 1354
Db 1997 ATGACAGCTGGCGGTGGGCAAGTTCGCGACGACCTTGGCGAAGCTCATCGAGCGTGT 2056
Qy 1355 TCGTCAAGTTCGTGAGCTGTGTCTCTTCAACAGGCGCTGTGTGGTTCGGCGGCGCG 1414
Db 2057 CCTTCAAGCTCTGTCAGCCAGCTGGAGGAGAGAGAGTTCGAGAGGCGCGCGAGCGT 2116
Qy 1415 TGGCGGCAAGCAGTACGCGCTGANTCAAGCGGACGCGCATCCCATTTGAAATCTATCT 1474
Db 2117 ACCTCGCTTCTGTGAAGGACTTCCCGCAGACGAGATCGCGACCTTGGCGCTCTACAACG 2176
Qy 1475 GCGCACCATGGACGCGGTGGCGGAGAACTCGACGTCGCGCAGCAGCAACTTCTTACT 1534
Db 2177 CGTCCGTCTGACTACTCAAGGCGGAGCGCTGGAATAGGCGCATCGAGGTGCGCAAGCGC 2236
Qy 1535 ARAATCCCTCACCGGCAAGTTCCTGGCGCAACTGCCCACTACTCTGGCGAGCGG 1594
Db 2237 TGTTCGCGGAGTACCCCGGCTCCAGACAGCTGCGGACTCCATCTAGCGAAACGCGAGG 2296
Qy 1595 CCTTGGCACCTCAAGAGTGGGTGGTGGACAACTGACCGTCTCCACCAACTTCTTCA 1654
Db 2297 CGCTGGAGGCAATTGGCGACTTCGAGGACGCGCGGCAACGTAAGGCGCTTACGTGGCG 2356
Qy 1655 TGGAGGAGCTCAAGCGCGGACCTACACCAAGTGTATTCGATGGACACGTCGATGGC 1714
Db 2357 GCTACGAGCGGAGCTCTGGGTGACAAAGGCAACGCAAGGCGGTGTGGCGGCAAGAGA 2416
Qy 1715 TGGATATGCCGTGGCGCAACGAGCTGCGCGAGTCCCTGGCCAAAGAGTGTGGCG---GG 1771
Db 2417 AGCGGGGCGGTGGAGCAACAAAGCCCGGTGTGTGAGAGTGGAGAGTCCAAAG 2476
Qy 1772 GCGCATCGTCACTGGCGTTCGCGCTCCCTCAGCCCGCCCTACGCGAGCTGATCCAGA 1831
Db 2477 GCGAGTGGCGCTCTTCAAGCGCGCCACCTACCGAGAGGCGCTGGGCGCAGGTGAAGCGG 2536
Qy 1832 AGCGGGGTTGAGCTGGGTGATCGCGCGCGCACTCAGGGCTATATGACCGCGTCA 1891
Db 2537 CGTTCGCAACCGCGAGCACTTACCTGACGTGTGGCGCGCGGCGAAGGACGCGCGACAA 2596
Qy 1892 ACATGTACAGCTCTTACATGCGCGCGCGGAGGCGCG 1932
Db 2597 TCCGCTCTCCATCATCTGACCTGACGCGCAAGAGCGCGCG 2637
```

RESULT 13

```
US-09-902-540-694
; Sequence 694, Application US/09902540
; Patent No. 683447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
```

```
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 694
; LENGTH: 5656
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
; US-09-902-540-694
```

```
Query Match 3.2%; Score 62.2; DB 3; Length 5656;
Best Local Similarity 42.3%; Pred. No. 0.0042;
Matches 474; Conservative 0; Mismatches 638; Indels 9; Gaps 2;
```

```
Qy 821 TCCTGTACAGCGCAGTCTGTGGAGGACCCCGAGCCGATATGGAGGTGATGAGATCAACC 880
Db 3573 TCCAGAGCGCAGAGTGTGGAGGCCAAGGACGACAAAGGCAAGCGGCAAGC 3632
Qy 881 CCAAGAGACAGGTGTGACCTGACTAGCGGCGGCTGCATGCGCTGAACTCTGTGTGC 940
Db 3633 CGGCGAAGTGGCTGCAGAAACGCGCCTACAAACGCGTGTGGCTTACGACAGGTCTCA 3692
Qy 941 AGGGGCGCGCCAGGTGTGTGCTGACTGCAACCCCGCGCAGTCGGCGCTTCTGAGC 1000
Db 3693 AGCGGCGGAGGCGCGGCGGAGGCAAGGCGGCGGTGGCTCGACATCAACAGA 3752
Qy 1001 TGAAGAAGGTGGCCATTTCAGCAGCTGGAGTTTGAAGGACGTGTGCGCAGCTGTTCGGCGAGG 1060
Db 3753 AGGCCACCATCCCAACGCTGAAGAGGCGCTGCTCGACGCGTGTGAGCGCTACTCTCAAGT 3812
Qy 1061 GCGTGACCGCGCATTTGAGGAGCTGTACGAGAGAGCTGGCGCGCTTCTGTGCGAAA 1120
Db 3813 ACGTGCCCAAGGTTGAGAGCGGTTGAGATCCCTTCAAGGCGGCGCAACATCTACTACC 3872
Qy 1121 CCAGCCACAACTTCTGCTCCAGCGCTCTGTGACTTCCAGACAGCGCTCTACTACCAGG 1180
Db 3873 GCCAACCACTTTCGACGAGCGGTGTGCGCTTTCAGCGAGATGCGCTCGGTACTCCCG 3932
Qy 1181 GCGGCATGGCAAGCTGTGTGGGTGTGTCAGTGCCTTGGCGCGGTGTGTGGGCA 1240
Db 3933 AGTACAAGTTCGAGAGCGGCGCGCGGAGATCTCCGCCAACCTCATCTCGACT 3992
Qy 1241 AGACCGTCAAGCGCTCGCCAAACGCGCCCAATATGGAGGAGCAGCGCGTCTGTGGACA 1300
Db 3993 CGTACCACTCTGCTCAGGACTACGCGAAGGTGAACGAGTGGCGCGCGCTTCTACGCCA 4052
Qy 1301 GCACATGC-----TCATCCACTTCGTGAAGAACGCGGCCAAAGCGCTGTGTGGCTGT 1354
Db 4053 ATGACAAAGTGGCGGTGGGCAAGTTCGCGACGACCTTGGCGAAGCTCATCGACAGTCTGT 4112
Qy 1355 TCGTCAAGTTCGTGAGCTGTGTCTTTCAACAAAGGCGGTGTGTGTGGTTCGGCGCGCGG 1414
Db 4113 CTTTCAAGCTCTGTCAGCCAGCTGGAGGAGAAAGGAGTTTCGAGAGGCGCGCGAGCGT 4172
Qy 1415 TGGCGGCAAGCAGTACGCGCTGANTCAAGCGGACGCGCATCCCATTTGAGAACTACATCG 1474
Db 4173 ACCTCGCTTCTGTGAAGGACTTCCCGCAGACGAGATTCGCGGCTCTCTACAAAG 4232
Qy 1475 GCGGCACCATGGAGCGGTGGCGGAGAACTCGACGTCGCGCAAGCAGAACTACTTCTACT 1534
Db 4233 CGTCCGTCTGACTACTCAAGCGCAAGCGCTTGATAGGCCATTCGAGGTGGCAAGCGCC 4292
Qy 1535 ACAACTGCTCTCACCGCAAGTTCCTGCGCGCAACTCTGCCCCACTACTCTCGCGAGCGG 1594
Db 4293 TGTTCGCGAGTACCCCGCGCTCCAAAGCAGCTGCGGACTCCATCTACGCGAAGCGGAGG 4352
Qy 1595 CTTTGGCCACCTCAAGAGTGGCGGTGTGAGAACCTGACGCTCCACCACTTCTTCA 1654
Db 4353 CGCTGAGGCGCAITTTGGCGACTTTCGAGGACGCGCGGCAACGTTAGGAGCGCTTACGTGCGG 4412
Qy 1655 TGGAGGAGCTCAAGCGCGCACCTTACACCAAGGTGATTCTGTGATGACACAGCTGACCTGGC 1714
Db 4413 GCTACGAGCGCAGCTCTGGGTGACAAAGGCAACGCCAAGGCGCGGTGTGGCGGCAAGAGA 4472
```

QY 1715 TGATATGCGCGGCAACGAGTGGCCGAGTGCCCTGGCCCAAGCAGGTTGCGCC---GG 1771
Db |||||
QY 4473 AGCGGGCGGGTGGAGCAACAAGCCCGCGTGGTGACAGAGTGGGACAGTCCAAAG 4532
Db |||||
QY 1772 GCGGCATCGTCATCTGGCGTCCGCTCCCTCAGCCCGCCCTAGCGGAGCTATCCAGA 1831
Db |||||
QY 4533 CGCAGGTGGCGCTCTCAACCGCGCCACCTACCGAGAGGCGCTGGGGCCAGGTGAAGCGCG 4592
Db |||||
QY 1832 AGCGGGCTTTCGACGTGGCTGTCATCGCGCGCCACTCAGGGGTATACATGACCGGGTCA 1891
Db |||||
QY 4593 CGCTGCGCAACCGGAGCACTACTGACGCTGTGGCGCGCGGAGGAGCGCCGACGAA 4652
Db |||||
QY 1892 ACATGTACAGCTCTTCTTACATGGCGCGCGGAGGCGCC 1932
Db |||||
QY 4653 TCGCCTCTCATCATCGACCTGACGGCCAAAGAGCGCGCC 4693
Db |||||

RESULT 14

US-09-902-540-6237
; Sequence 6237, Application US/09902540
; Patent No. 6833447

GENERAL INFORMATION:

; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 6237
; LENGTH: 981
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-6237

Query Match 3.2%; Score 61.8; DB 3; Length 981;
Best Local Similarity 49.8%; Pred. No. 0.0038;
Matches 156; Conservative 0; Mismatches 157; Indels 0; Gaps 0;

QY 1443 GCGGACGGCATCCCATTTGAGAACTACATCGCGCGCACCATGGACGGCGTGGCGGAGAA 1502
Db |||||
QY 246 GGACGGCAGCGGTACTCTCAATCAGTCTCTGTGGTGGCAAGCGCGCGCCGAGCC 305
Db |||||
QY 1503 CTGCGACGTGGCAAGCAGAACTACTTCTACTCAACTGCTCCTCAGCGGCAAGTTCCTGG 1562
Db |||||
QY 306 CCGGAGACGTTGAAGACGACCGCTGTGTACAGGGCGGCTCCGGGAGCTTCTCTGGC 365
Db |||||
QY 1563 CGACAACTGCGCCACCTACTCTGCGCGGCGGCTTGGCCACCTCAAGAGTGGCGTGGT 1622
Db |||||
QY 366 GCCACCGCGCAATCCCCCTGGCGACGAGGGCTGGGGCTTGGATTTTCGAAGGCGAGGT 425
Db |||||
QY 1623 GGACAACTGACCGTCTCCACCAACTTCTTCTATGGAGAGCTCAAAAGCGCGCACCTACAC 1682
Db |||||
QY 426 CTGCGCATCTGGCGGACAGCGCCCGAGGCAAGAGGCGGAGAACGCGGCGCGCACGT 485
Db |||||
QY 1683 CAAGGTGATTTCTGATGACCACTGTGATGATGATGATGATGATGATGATGATGATGATG 1742
Db |||||
QY 486 CAAGCTGCTGATGCTGGCCAAATGACGCTGTCCCTGGCGCAACCTCATCCCCAACGAGCTGGC 545
Db |||||
QY 1743 CGAGTGCCTGGCC 1755
Db |||||
QY 546 CAAGGGCTTCGAC 558
Db |||||

RESULT 15

US-09-902-540-1047
; Sequence 1047, Application US/09902540
; Patent No. 6833447

GENERAL INFORMATION:

; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1047
; LENGTH: 1332
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-1047

Query Match 3.2%; Score 61.8; DB 3; Length 1332;
Best Local Similarity 49.8%; Pred. No. 0.0058;
Matches 156; Conservative 0; Mismatches 157; Indels 0; Gaps 0;

QY 1443 GCGGACGGCATCCCATTTGAGAACTACATCGCGCGCACCATGGACGGCGTGGCGGAGAA 1502
Db |||||
QY 11288 GGACGGCAGCGGTACTCTCAATCAGTCTCTGTGGTGGCAAGCGCGCGCCGAGCC 11347
Db |||||
QY 1503 CTGCGACGTGGCAAGCAGAACTACTTCTACTCAACTGCTCCTCACCGGCAAGTTCCTGG 1562
Db |||||
QY 11348 CCGGAGACGTTGAAGACGACGCCGCTGTGTACAGGGCGGCTCCGGGAGCTTCTCTGGC 11407
Db |||||
QY 1563 CGACAACTGCGCCACCTACTCTGCGCGAGGCGGCTTGGCCACCTCAAGAGTGGCGTGGT 1622
Db |||||
QY 11408 GCCACCGCGCACATCCCCCTGGCCGACGAGGCTTGGGGCTTGGATTTTCGAAGGCGAGGT 11467
Db |||||
QY 1623 GGACAACTGACCGCTCTCCACCAACTTCTTCTATGGAGGAGCTCAAAAGCGCGCACCTACAC 1682
Db |||||
QY 11468 CTGCGCATCTGGGCGACACGCCCCAGGGCACGAAGCGGAGAACGCGCGCGCGCACGT 11527
Db |||||
QY 1683 CAAGGTGATTTCTGATGACCACTGCTGAGCTGGCTGGTATGATGATGATGATGATGATGATG 1742
Db |||||
QY 11528 CAAGCTGCTGATGCTGGCCAAATGACGCTGTCCCTGGCGCAACCTCATCCCCAACGAGCTGGC 11587
Db |||||
QY 1743 CGAGTGCCTGGCC 1755
Db |||||
QY 11588 CAAGGGCTTCGAC 11600
Db |||||

Search completed: March 14, 2006, 01:18:31
Job time : 371 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 14, 2006, 00:48:26 ; Search time 1628 Seconds
(without alignments)
9889.729 Million cell updates/sec

Title: US-10-620-914-44

Perfect score: 1947
Sequence: 1 atggggctcggtcgtagcgg.....gagccgaagaggaactaa 1947

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_NA_Main:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1947	100.0	1947	7	US-10-620-914-44
2	645	33.1	5242	7	US-10-620-914-43
3	100.2	5.1	1252	5	US-10-118-495-1
4	100.2	5.1	1252	7	US-10-620-914-1
5	76.8	3.9	2259	8	US-10-741-849-6021
6	73	3.7	2727	7	US-10-620-914-49
7	73	3.7	3427	7	US-10-620-914-48
8	72.4	3.7	137560	8	US-10-481-112-1
9	70	3.6	1251	5	US-10-118-495-32
10	70	3.6	1251	7	US-10-620-914-32
11	69.6	3.6	1248	5	US-10-118-495-22
12	69.6	3.6	1248	7	US-10-620-914-22
13	67.8	3.5	1104	6	US-10-204-434A-12
14	67.8	3.5	6798	3	US-09-918-740-57
15	67.8	3.5	6798	6	US-10-204-434A-13
16	67.8	3.5	6798	8	US-10-835-516-57
17	67.8	3.5	6798	10	US-11-053-541-57
18	67.8	3.5	8077	3	US-09-918-740-63
19	67.8	3.5	8077	6	US-10-835-516-63
20	67.8	3.5	8077	10	US-11-053-541-63
21	67.6	3.5	5858	8	US-10-488-056-14
22	66.6	3.4	1626	7	US-10-392-041-3
23	66.6	3.4	1626	7	US-10-394-763-3

ALIGNMENTS

RESULT 1
US-10-620-914-44
; Sequence 44, Application US/10620914
; Publication No. US20040093639A1
; GENERAL INFORMATION:
; APPLICANT: Benning, Christoph
; APPLICANT: Riekhof, Wayne
; APPLICANT: Klug, Rouven
; TITLE OF INVENTION: Compositions and Methods for the Production of Betaine Lipids
; FILE REFERENCE: MSU-07769
; CURRENT APPLICATION NUMBER: US/10/620,914
; CURRENT FILING DATE: 2003-07-16
; PRIOR APPLICATION NUMBER: 10/118,495
; PRIOR FILING DATE: 2002-04-08
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 44
; LENGTH: 1947
; TYPE: DNA
; ORGANISM: Chlamydomonas reinhardtii
US-10-620-914-44

Query Match	Best Local Similarity	Score 1947;	DB 7;	Length 1947;
Matches 1947;	Conservative	0;	Mismatches	0;
Indels	0;	Gaps	0;	
Qy	1	ATGGGGTCTGGTCTGTCAGCGCGCGCTCGAGCTACACCAAGAACTTCTCCCTGGAG	60	
Db	1	ATGGGGTCTGGTCTGTCAGCGCGCGCTCGAGCTACACCAAGAACTTCTCCCTGGAG	60	
Qy	61	AAGCTCAAGCTCAGCAGCATGAAGATGACCTGACCGCTTCTGCGCATATGTGTTCCGC	120	
Db	61	AAGCTCAAGCTCAGCAGCATGAAGATGACCTGACCGCTTCTGCGCATATGTGTTCCGC	120	
Qy	121	AGCAAGAAAGGGGATGATCAGCTGCTCGCTGAGAGCTTCTACGGGCCCCAGCGCGCT	180	
Db	121	AGCAAGAAAGGGGATGATCAGCTGCTCGCTGAGAGCTTCTACGGGCCCCAGCGCGCT	180	
Qy	181	GCCTTTGCTGCCCGCTGCGCGAGCGCTCGAACCTCATCTGGGTGACCTGGGTGGTGGC	240	
Db	181	GCCTTTGCTGCCCGCTGCGCGAGCGCTCGAACCTCATCTGGGTGACCTGGGTGGTGGC	240	
Qy	241	ACTGGGGAGAAATGTCGATATGATGGCTGATTACATGACCTGCGGAAGTTCAAGTCCATC	300	
Db	241	ACTGGGGAGAAATGTCGATATGATGGCTGATTACATGACCTGCGGAAGTTCAAGTCCATC	300	

1165 GGCTGTACTACAGGCGGCATGGCAAGCTGTGCTGGGTGCTGCAAGTCCCTGGCGTG 1224
1166 |
3348 GGCTGTACTACAGGCGGCATGGCAAGCTGTGCTGGGTGCTGCAAGTCCCTGGCGTG 3407
1225 GTCTGGGACTGGGCAAGCGCTGAGCGCTCGCCACCGGCCCAATGGAGGAGCAG 1284
3408 GTCTGGGACTGGGCAAGCGCTGAGCGCTCGCCACCGGCCCAATGGAGGAGCAG 3467
1285 CGCGCTGTGGGACAGCAATGCTCATCACTTCGTGAAGAAAGCGGCCCAAGCGCTG 1344
3468 CGCGCTGTGGGACAGCAATGCTCATCACTTCGTGAAGAAAGCGGCCCAAGCGCTG 3527
1345 GTGTGGCTGTGTCAGTTGCTGAGCGCTGCTTCCTTCAAGGCGCTGCTGCTGCTTC 1404
3528 GTGTGGCTGTGTCAGTTGCTGAGCGCTGCTTCCTTCAAGGCGCTGCTGCTGCTTC 3587
1405 GCGGCGGCGCTGCGGCGGACAGTACCGCTCATCAAGCGGCGGCGCATCCCATGAG 1464
3588 GCGGCGGCGCTGCGGCGGACAGTACCGCTCATCAAGCGGCGGCGCATCCCATGAG 3647
1465 AACTATACATCGCGGCGCATCGGCGGAGAACTCGCAAGCGGCGGCGCATCGGCGG 1524
3648 AACTATACATCGCGGCGCATCGGCGGAGAACTCGCAAGCGGCGGCGCATCGGCGG 3707
1525 TACTTCTACTACAACTGCTTACCGGCAAGTTCCTGCGGCAAACTGCGGCGGCGG 1584
3708 TACTTCTACTACAACTGCTTACCGGCAAGTTCCTGCGGCAAACTGCGGCGGCGG 3767
1585 GCGGAGCGGCGCTTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1644
3768 GCGGAGCGGCGCTTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3827
1645 AACTTCTTATGAGGAGCTCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1689
3828 AACTTCTTATGAGGAGCTCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3872

RESULT 3

US-10-118-495-1
; Sequence 1, Application US/10118495
; Publication No. US20030074688A1
; GENERAL INFORMATION:
; APPLICANT: Benning, Christoph
; APPLICANT: Riekhof, Wayne
; APPLICANT: Klug, Rouven
; TITLE OF INVENTION: Compositions and Methods for the Production of Betaine Lipids
; FILE REFERENCE: MSU-06897
; CURRENT APPLICATION NUMBER: US/10/118,495
; CURRENT FILING DATE: 2002-04-08
; PRIOR APPLICATION NUMBER: 60/283,812
; PRIOR FILING DATE: 2001-04-13
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1252
; TYPE: DNA
; ORGANISM: Rhodobacter sphaeroides
US-10-118-495-1

Query Match 5.1%; Score 100.2; DB 5; Length 1252;
Best Local Similarity 50.9%; Pred. No. 2e-16;
Matches 299; Conservative 0; Mismatches 273; Indels 15; Gaps 2;
809 TCCGCGCCACCTTCTGTACACGAGTCTGTGGGAGGACCGGCGGATATGAGGTGA 868
122 TCTTCCACGCGCTCGTCTATCCGAGATCTGGGAGGATCCGGCGGTGGACATGGCGGCC 181
869 TGGAGATCAACCCCAAGGACACGGTGTCTGACCTCTGACCTAGCGGCGGCTGCAATGCCCTGA 928
182 TCGCCATCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 241
929 ACCTGTGTGTCAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 988

242 CCTATCTACAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 301
989 CGCTTCTGAGAGTGAAGAAGTGGCCATTCAGCAGCTG---GAGTTTGGAGACGTGGGC 1045
302 CGCTGGGCGGCTGAAGCTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 361
1046 AGCTGTTCGCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1105
362 ATCTTCTGCTCGCGCAGACCTGCCCGGCAATGCGGCGCTTACGACCGGCGCACATCGCG 421
1106 CTTTCTGTTCGCAACACGAGCCAACTTCTGTGTCGAGGCGGCGGCGGCGGCGGCGG 1165
422 CCGCGCTCGACGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 481
1166 GCCTGTACTACAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG --- 1222
482 TCAGAGCTGTTCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 541
1223 -----TGTGCTGGGACTGGGCAAGACCGTCAAGCGGCGGCGGCGGCGGCGGCGG 1273
542 CCATACGCTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 601
1274 TGGAGAGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1333
602 TCGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 661
1334 CCAAGCGGCTGTGCTGCTCAAGTTCGTGAGCGGCGGCGGCGGCGGCGGCGGCGG 1380
662 TCGAGCGGCTTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 708

RESULT 4

US-10-620-914-1
; Sequence 1, Application US/10620914
; Publication No. US20040093639A1
; GENERAL INFORMATION:
; APPLICANT: Benning, Christoph
; APPLICANT: Riekhof, Wayne
; APPLICANT: Klug, Rouven
; TITLE OF INVENTION: Compositions and Methods for the Production of Betaine Lipids
; FILE REFERENCE: MSU-07769
; CURRENT APPLICATION NUMBER: US/10/620,914
; CURRENT FILING DATE: 2003-07-16
; PRIOR APPLICATION NUMBER: 10/118,495
; PRIOR FILING DATE: 2002-04-08
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 1252
; TYPE: DNA
; ORGANISM: Rhodobacter sphaeroides
US-10-620-914-1

Query Match 5.1%; Score 100.2; DB 7; Length 1252;
Best Local Similarity 50.9%; Pred. No. 2e-16;
Matches 299; Conservative 0; Mismatches 273; Indels 15; Gaps 2;
809 TCCGCGCCACCTTCTGTACACGAGTCTGTGGGAGGACCGGCGGATATGAGGTGA 868
122 TCTTCCACGCGCTCGTCTATCCGAGATCTGGGAGGATCCGGCGGTGGACATGGCGGCC 181
869 TGGAGATCAACCCCAAGGACACGGTGTCTGACCTCTGACCTAGCGGCGGCTGCAATGCCCTGA 928
182 TCGCCATCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 241
929 ACCTGTGTGTCAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 988
242 CCTATCTACAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 301
989 CGCTTCTGAGAGTGAAGAAGTGGCCATTCAGCAGCTG---GAGTTTGGAGACGTGGGC 1045
302 CGCTGGGCGGCTGAAGCTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 361

QY 1046 AGCTGTTGCGAGGGCGTGCACCCGCGCATTGAGGAGCTGTACGAGAAAGCTGCGGC 1105
DB 362 ATCTCTTTGCGTGGCGAGACTGCGCGCAATGCGGCCCTCTACGACCGCCACATCGCGC 421
QY 1106 CCTTCTGTGCAAAACAGCAGCAAACTTCTGGTCCAAAGCGCCCTCTGTACTTCCAGCACG 1165
DB 422 CCGGCTCGAGCGCGCGGAGCGCGCTACTTGGGAGCGCGCAGCCCTTTCGGCGCGGCA 481
QY 1166 GCTGTACTACAGGGCGGATGGGCAAGCTGTGCTGGGTGCTGAGTGTGCTGGCGG--- 1222
DB 482 TCCAGCTGTTGAGCGCGGCTTCTTACCGGCAAGCTGCTTCCGCGCGCTTTCATCGCGCGG 541
QY 1223 -----TGGTGTGGGACTGGGCAAGACCGTCAAGCGCCCTCGCCAGCGCGCCACAA 1273
DB 542 CCATAGCTTGGCGGCGCGCGGACCGACCTGCGGGGCTTCTGAGTGTCCGACA 601
QY 1274 TGGAGGAGCAGCGCGCTGTGCGGACAGCAACATGTCTATCCACTTGTGTAAGAAGCGGC 1333
DB 602 TCGAGGCGCAGCGAGCTTCTTACGCCCATATCGGGCGCTTTCGAGGCGCGCGTGG 661
QY 1334 CCAAGCGCTGTGTGCTGTGCTCAAGTTCGTGAGCCCTGTGTGCTC 1380
DB 662 TGCAGGCGCTCGCGGCGCGCGCGCGCTTTCGGGCTGGGGATC 708

RESULT 5

US-10-741-849-6021
; Sequence 6021, Application US/10741849
; Publication No. US20050019931A1

GENERAL INFORMATION:

; APPLICANT: Roemer, Terry
; APPLICANT: Jiang, Bo
; APPLICANT: Boone, Charles
; APPLICANT: Bussey, Howard
; TITLE OF INVENTION: Nucleic Acids Encoding Anti-fungal Drug Targets and Methods of
; TITLE OF INVENTION: Use
; FILE REFERENCE: 10182-023-999
; CURRENT APPLICATION NUMBER: US/10741,849
; CURRENT FILING DATE: 2003-12-19
; PRIOR APPLICATION NUMBER: US 60/434,832
; PRIOR FILING DATE: 2002-12-19
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6021
; LENGTH: 2259
; TYPE: DNA
; ORGANISM: Candida albicans
US-10-741-849-6021

Query Match 3.9%; Score 76.8; DB 8; Length 2259;
Best Local Similarity 54.3%; Pred. No. 4.2e-10;
Matches 178; Conservative 0; Mismatches 147; Indels 3; Gaps 1;
QY 212 ACCTCATCTGGTGTGACTGGGTGGTGGCACTGGGAGAAATGTCGATATGATGCTCAT 271
DB 323 ATCTATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 382
QY 272 ACATCGACCT---GGCGAAGTTCAAGTCCATCTACGTGGTGGAGCTGTGCGACCTGCTGT 328
DB 383 TTAGTAAATATCTGAAAACCTTTAAAGCTGTTTATTTGGTGTGATCTTTTCCCATCTTTGT 442
QY 329 GCGAGTGGCCAAAGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGG 388
DB 443 GTGAAGTGTGTAAGGCAAGATTGTAAGCCCATGTAATGGAACAAATGTTTCATGATTAGTTG 502
QY 389 CCGACGCTTGGCAATTTGGCGCCCTCGAGGCGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 448
DB 503 CTGATGCTGTGATTTTACATGATGATGATGATGATGATGATGATGATGATGATGAT 562
QY 449 CGCTCAGATGATTTCCACCGTTTCCACACGTCGACGAGGCTTGTCTGACCTGCTGCC 508
DB 563 CATTGTGATGATCCCAACTTTTCAATGCTGTATCGATTAATGCTGTTTCTTAAATTAGATA 622

QY 509 AAGACGCGCTGGTGGCGCTTTCGCGACTT 536
DB 623 TGAAGGTATTATTGCCACTGTGATTT 650

RESULT 6

US-10-620-914-49
; Sequence 49, Application US/10620914
; Publication No. US20040093639A1
; GENERAL INFORMATION:
; APPLICANT: Banning, Christoph
; APPLICANT: Riekhof, Wayne
; APPLICANT: Klug, Rouven
; TITLE OF INVENTION: Compositions and Methods for the Production of Betaine Lipids
; FILE REFERENCE: MSU-07769
; CURRENT APPLICATION NUMBER: US/10/620,914
; CURRENT FILING DATE: 2003-07-16
; PRIOR APPLICATION NUMBER: 10/118,495
; PRIOR FILING DATE: 2002-04-08
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 49
; LENGTH: 2727
; TYPE: DNA
; ORGANISM: Neurospora crassa
US-10-620-914-49

Query Match 3.7%; Score 73; DB 7; Length 2727;
Best Local Similarity 51.0%; Pred. No. 4.5e-09;
Matches 172; Conservative 0; Mismatches 165; Indels 0; Gaps 0;

QY 779 ACCGAGCGCGTGGAGCGCGCCCATGTTCCCGCCACTTTCCTGTGTACACGAGTCGT 838
DB 1448 ACACGATCAACTCCGAGACACACCAAGTTCATGACGAGTACATCTAGCCTTTACCT 1507.
QY 839 GCGAGGACCCCGAGCGCGGATATGAGGTGATGAGATCAACCCCAAGGACACCGTGTGA 898
DB 1508 GCGAGACTCGCGGTCGACAGAGAACTCTTAACCTCGGGCCGACGAGCTGCTCTAG 1567
QY 899 CCTGACTAGCGCGGTGCAATGCCCTGACCTGTGTCGAGGGGCGCGGAGTGG 958
DB 1568 CCATCAGCAGCGCGCGCAACATTTCTTCTACCTGATGACAGAGTCCCGCTGCGCTGC 1627
QY 959 TGTGCGGTGAGTCAACCCCGCGCGCTTCTGAGAGTCAAGAGGTGCGCAATTC 1018
DB 1628 ACGCCATCGACTTAACCCGAGCCCAAAACCACTGTTGAACTCAAAGTGCCTCTTTA 1687
QY 1019 AGCAGCTGGAGTTGAGGACGCTGTGGCAGCTGTTCGCGAGGCGGTGACACCGCGCAT 1078
DB 1688 CGACTCTGGATTACCCCGAGCTCTGGAAGATCTTCGGTGGAGGCAACACACCCCGACTTC 1747
QY 1079 AGGAGCTGTACGAGAGAGAGTGGCGCCCTTCTCTGTC 1115
DB 1748 GCTCACTGCTCATCTCCAAACTCTCCCTCACCTCTC 1784

RESULT 7

US-10-620-914-48
; Sequence 48, Application US/10620914
; Publication No. US20040093639A1
; GENERAL INFORMATION:
; APPLICANT: Banning, Christoph
; APPLICANT: Riekhof, Wayne
; APPLICANT: Klug, Rouven
; TITLE OF INVENTION: Compositions and Methods for the Production of Betaine Lipids
; FILE REFERENCE: MSU-07769
; CURRENT APPLICATION NUMBER: US/10/620,914
; CURRENT FILING DATE: 2003-07-16
; PRIOR APPLICATION NUMBER: 10/118,495
; PRIOR FILING DATE: 2002-04-08
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.2

```
; SEQ ID NO 48
; LENGTH: 3427
; TYPE: DNA
; ORGANISM: Neurospora crassa
US-10-620-914-48

Query Match      3.7%; Score 73; DB 7; Length 3427;
Best Local Similarity 51.0%; Pred. No. 4.8e-09;
Matches 172; Conservative 0; Mismatches 165; Indels 0; Gaps 0;

QY 779 ACAGAGAGCGCTGGAGCGCGCCCATGTTCCCGCCACCTTCCTGTACAGCGAGTGGT 838
Db 1906 ACAGAGTAACTCCCGAAGCACCACGATTCATGACGAGTACATCTACGCTTTACCT 1965

QY 839 GGAGAGACCCCGAGCCGGATATGGAGGTGATGGAGATCAACCCCAAGGACACGGTGTGA 898
Db 1966 GGGAAGACTCGCGCTGCGACAGAGAACTCTTAACCTCGGGCCGCGAGCGTCTCTAG 2025

QY 899 CCTGACTAGCGGGCTGCAATGCCCTGAACTCTGTGTCAGGGGGCCGCGAGGTGG 958
Db 2026 CCATCACCAGCGCGCGCAACATTTCTTCTACCTGATGACAGAGTCCCGCTCGCGTGC 2085

QY 959 TGTGCGTGGAGTCAACCCCGCGAGTCGCGCTTCTGGAGCTGAAGAAGTGGCCATTC 1018
Db 2086 ACCCATCGACCTAAACCCAGCCAAACACCACTGCTTGAACCTCAAGTCCCTCTTTA 2145

QY 1019 AGCAGCTGGAGTTTGAGGAGCTGTGGCAGCTGTTCGGCGAGGGCGTGCACCCGCGCATTG 1078
Db 2146 CGACTCTGGATTACCCCGAGCTGTGGAAGATCTTCGGTGAGGGCAACACCCCGACTTTC 2205

QY 1079 AGGAGCTGTACGAGAGAGAGTGGCGCCCTTCTCTGTC 1115
Db 2206 GTCACCTGCTCATCTCCAAACTCTCCCTCACCTCTC 2242

RESULT 8
US-10-481-112-1
; Sequence 1, Application US/10481112
; Publication No. US20040235721A1
; GENERAL INFORMATION:
; APPLICANT: WEBER, Olaf
; APPLICANT: FRIEDERICH, Sonja, Maria
; APPLICANT: SIEGLING, Angela
; APPLICANT: SCHLAPP, Tobias
; APPLICANT: MERCER, Andrew, Allan
; APPLICANT: FLEMING, Stephen, Bruce
; TITLE OF INVENTION: Recombinant proteins of Parapoxvirus ovis and pharmaceutical comp
; FILE REFERENCE: Lea 35228
; CURRENT APPLICATION NUMBER: US/10/481,112
; CURRENT FILING DATE: 2003-12-12
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 137560
; TYPE: DNA
; ORGANISM: Parapoxvirus ovis NZ2
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3)...(539)
; OTHER INFORMATION: ORF: L1
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (781)...(449)
; OTHER INFORMATION: ORF: L2r
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1933)...(1664)
; OTHER INFORMATION: ORF: L3r
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3269)...(2790)
; OTHER INFORMATION: ORF: L4r
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2799)...(3851)
; OTHER INFORMATION: ORF: L5
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2962)...(3753)
; OTHER INFORMATION: ORF: L6
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3784)...(3122)
; OTHER INFORMATION: ORF: L7r
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (4341)...(4129)
; OTHER INFORMATION: ORF: L8r
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (4904)...(4428)
; OTHER INFORMATION: ORF: L9r
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (6517)...(4970)
; OTHER INFORMATION: ORF: L1r
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (8042)...(6684)
; OTHER INFORMATION: ORF: L2r
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (9989)...(8070)
; OTHER INFORMATION: ORF: L3r
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (11195)...(10062)
; OTHER INFORMATION: ORF: L4r
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (11493)...(11227)
; OTHER INFORMATION: ORF: L5r
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (11802)...(12038)
; OTHER INFORMATION: ORF: L6
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (12358)...(12080)
; OTHER INFORMATION: ORF: L7r
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (13980)...(12364)
; OTHER INFORMATION: ORF: L8r
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (14826)...(14053)
; OTHER INFORMATION: ORF: L9r
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (15080)...(15394)
; OTHER INFORMATION: ORF: L10
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (16838)...(15423)
; OTHER INFORMATION: ORF: L11r
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (19021)...(16847)
; OTHER INFORMATION: ORF: L12r
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (19704)...(19156)
; OTHER INFORMATION: ORF: L13r
; FEATURE:
```

```
/ NAME/KEY: CDS
/ LOCATION: (20314)...(19736)
/ OTHER INFORMATION: ORF: 14r
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (20401)...(22101)
/ OTHER INFORMATION: ORF: 15
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (22125)...(22940)
/ OTHER INFORMATION: ORF: 16
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (23003)...(23866)
/ OTHER INFORMATION: ORF: 17
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (26908)...(23873)
/ OTHER INFORMATION: ORF: 18r
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (26926)...(27213)
/ OTHER INFORMATION: ORF: 19
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (27626)...(27216)
/ OTHER INFORMATION: ORF: 20r
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (29754)...(27616)
/ OTHER INFORMATION: ORF: 21r
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (32217)...(29800)
/ OTHER INFORMATION: ORF: 22r
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (33380)...(32418)
/ OTHER INFORMATION: ORF: 23r
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (33602)...(33393)
/ OTHER INFORMATION: ORF: 24r
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (34466)...(33612)
/ OTHER INFORMATION: ORF: 25r
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (34735)...(34502)
/ OTHER INFORMATION: ORF: 26r
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (35905)...(34739)
/ OTHER INFORMATION: ORF: 27r
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (37194)...(35905)
/ OTHER INFORMATION: ORF: 28r
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (37200)...(39248)
/ OTHER INFORMATION: ORF: 29
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (41037)...(39229)
/ OTHER INFORMATION: ORF: 30r
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (41374)...(42066)
/ OTHER INFORMATION: ORF: 31
/ FEATURE:
/ NAME/KEY: CDS
```

```
/ LOCATION: (42336)...(41731)
/ OTHER INFORMATION: ORF: 32r
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (42407)...(41997)
/ OTHER INFORMATION: ORF: 33r
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (42410)...(43765)
/ OTHER INFORMATION: ORF: 34
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (43770)...(43958)
/ OTHER INFORMATION: ORF: 35
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (43980)...(44534)
/ OTHER INFORMATION: ORF: 36
```

Query Match 3.7%; Score 72.4; DB 8; Length 137560;
Best Local Similarity 42.9%; Pred. No. 1.5e-08;
Matches 461; Conservative 0; Mismatches 611; Indels 2; Gaps 2;

Qy	618	CGACAAATTGACATCGGCGCCCGAGCGCGCCCTTACCTGGAGCAGAGCTGGAGCCGCT	677
Db	20613	CGAGGTCTTCGACGCGCTCTCAAGCGCGCTTCGACGCGCCCTCAGCTCTTTGTGCGGC	20672
Qy	678	GTGGGAGCAGAACACACAGGTTTCGATCCCTACGTGCGGTGGCTGCGGCCCTACTA	737
Db	20673	GCTCTCGGCGCACTTCGAGGCCCTTCGTCGCGCATCCGGGCTTACCGCGGACCTCGC	20732
Qy	738	CGTGTGATTTGGCGGCTGCGCAGCGTTGGCCACGCGCTTCACGAGGAGCGCGTGGAGCG	797
Db	20733	GGAGCTGTGGCGCAGCGCGGCTTCATGGAGTGGCTGCGCGCGCGGAGCTCGCGGA	20792
Qy	798	GCCGCCCATGTTCCCGCCACCTTCTGTACAGCGAGTCTGGGAGGACCCCGAGCCGGA	857
Db	20793	GGTCATTGGCGTGAACACACATCGCGCAACCGCTCTTCGCGGACGCGGAGCCGCT	20852
Qy	858	TATGGAGGTGATGGAGATCAACCCCAAGGACAGGTCTGACCTGACTAGCGCGGCTG	917
Db	20853	GCGGACGCGGAGCTCATTTTCGCAAGACCTTCGCAAGACGAGTTCGCGCGGCTCAA	20912
Qy	918	CAATGCCCTGAACCTGCTGTGTCAGGGGCGCGCCAGGTGCTGCTGGTGGAGTGCACACCC	977
Db	20913	GCGCTCGCGCTGTGGGCTGTGCTGGGCTTCTCTGTAAGAAGACCTTTGGCGG	20972
Qy	978	CGCGCAGTCCGCGCTTCTGGAGCTGAAGAGGTGGCCATTCAGCAGCTGGAGTTGAGGA	1037
Db	20973	CGAGTACGCGACAAACGACCGCCAGGACCTGTTTACGCTGCTGCAGAAAGCGCGGGCC	21032
Qy	1038	CGTGTGGCAGCTGTTGGCGAGGGCGTGACCCGCGCATTTGAGGAGTGTACGAGAAGAA	1097
Db	21033	CGTGGCGCA-CAGCGCGCTCACAGAGAGCATCCGCGAGTACTCTTCCCGGAGACAGGC	21091
Qy	1098	GCTGGCGCCCTTCTGTGCAAAACACGACCACTTCTGTGCTCAAGCGCCT-CTGGTACT	1156
Db	21092	CCAGCCACTGGGTCTGGCTGAACGCGCGCTGGCGGACGCGAGAGTGTACCGGACC	21151
Qy	1157	TCCAGCACGGCCTGTACTACAGGGCGGCAATGGGCAAGCTGTGCTGGGTGCTGAGTGCC	1216
Db	21152	GGCCCGCGCGCAGCTCTACGAGCGCTGCTCAGTACGCGTACTCAGAGGTCAAGCAGG	21211
Qy	1217	TGGCGTGTGGTGGGACTGGGCAAGACCGTCAAGCGCTTCGCCAACCGCGCCCAATGG	1276
Db	21212	GGCGCGTGAACGCGCAACACGCTCAAGTCTGTGTACCGGCTCGAGAGCAACCCGACATCA	21271
Qy	1277	AGGAGCAGCGCGCTCTGTGGGACAGCAACATGCTCATCTCCATCTCGTGAAGAACGGGCCA	1336
Db	21272	AGGTCTGTGCTGCACCTCATCTACGCGTGCCTGCGGACATCGTGGCGCTGTGGACT	21331
Qy	1337	AGCGCTGGTGTGCTGTTCTGTTCAAGTTCGTGAGCCTGGTGTCTTCAACAGGCGCGTGC	1396

Db 21332 CCAGCAACGAGGAGTGCGGAGGACTTCTGTGAGTCTGTACCGAGAACTTCGTGACG 21391
QY 1397 TGTGGTTTCGGCGCGCGTCCCGGCAAGCAGTACGCGCTGTATCAAGCGGAGCGCATCC 1456
Db 21392 GACGCACTTACCTCGAGCGCGCTTCGCGAGCACTTTCGCGTGTGCGCGCG 21451
QY 1457 CATTGAGAACTACATCGCGGCAACCATGAGCGGCTGGCGGAGAACTTCGACGTGGCA 1516
Db 21452 TCATCCGCACTTCTTCGAGCGCGGAGCGCATCGCGAGGCTTCAGCGGAGCGCGCG 21511
QY 1517 AGCAGAACTACTTCTACTCAACTGCTCTACCGGCAAGTTCCTGCGGCAACTGCCCCA 1576
Db 21512 TCGAGAGCGCTTCACGGAATGAGCTCAACACGCTTCATGTGCGCACTTATTCAG 21571
QY 1577 CTTACTTCGCGAGCGCGCTTCGCCACCCCTCAAGAGTGGCGTGGGCAACCTGACCG 1636
Db 21572 ACTCGTGGACCCGAGCTCGCGCGCGGAGCGCGGCTCGCACTGCGGCTGCACACG 21631
QY 1637 TCTCCACCACTTCTTCATGAGGAGCTCAAGGCGGCACTTACCAAGGTGA 1690
Db 21632 AGGACTCCGACTTCTCATCCGGAGGTACAACTACCTTCTTCTCAGCGAGA 21685

RESULT 9

US-10-118-495-32
; Query Match 3.6%; Score 70; DB 5; Length 1251;
; Best Local Similarity 52.8%; Pred. No. 2.5e-08;
; Matches 151; Conservative 0; Mismatches 135; Indels 0; Gaps 0;
; Publication No. US20030074688A1
; GENERAL INFORMATION:
; APPLICANT: Benning, Christoph
; APPLICANT: Riekhof, Wayne
; APPLICANT: Klug, Rouven
; TITLE OF INVENTION: Compositions and Methods for the Production of Betaine Lipids
; FILE REFERENCE: MSU-06897
; CURRENT APPLICATION NUMBER: US/10/118,495
; CURRENT FILING DATE: 2002-04-08
; PRIOR APPLICATION NUMBER: 60/283,812
; PRIOR FILING DATE: 2001-04-13
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 32
; TYPE: DNA
; ORGANISM: Sinorhizobium meliloti
US-10-118-495-32

QY 806 TGTTCGCGCCACCTTCTGTACACGAGTCTGTGGGAGGAGCCCGGAGCGGATATGGAGG 865
Db 116 TGCTCTTTCCGAGCTCGTCTACCGCAGATCTGGGAGGAGCCGATTTGCGACATGGAAG 175
QY 866 TGATGAGATCAACCCCAAGGACACGGTGTGACCTGTAGCTAGCGGCGGTGCAATGCC 925
Db 176 CGATGAGATCCGCTCCCGGACATCGGATCGTACGATCGGTTCCGCGGCTGCAACATGC 235
QY 926 TGAACCTGTGTGAGGCGCGGCGGAGTGTGTGCTGCTGAGTGTGAGGAGTGTGGC 985
Db 236 TGACCTATCTCTCCGCGAGCTTCGCGGATAGACGTTGCGATCTCAACCCCATCA 295
QY 986 CGCGCTTCTGAGCTGGAAGAGTGGCAATTCAGACGCTGGAGTTTGGAGAGTGTGGC 1045
Db 296 TCGCGCTCAACCGCTGAGCTGTCTGCTTCCGCACTTCGCGGAGCCGACAGGAGCTGG 355
QY 1046 AGCTGTTCGCGAGGCGGTGCACCCCGCATTTAGGAGTGTGTACGA 1091
Db 356 TGGGTTCTCTCGCGTGAAGGTACGCGCAAGTGGCAAGTGGCAAGGCGCTTACCAAC 70

RESULT 10

US-10-620-914-32
; Sequence 32, Application US/10620914

; Publication No. US20040093639A1
; GENERAL INFORMATION:
; APPLICANT: Benning, Christoph
; APPLICANT: Riekhof, Wayne
; APPLICANT: Klug, Rouven
; TITLE OF INVENTION: Compositions and Methods for the Production of Betaine Lipids
; FILE REFERENCE: MSU-07769
; CURRENT APPLICATION NUMBER: US/10/620,914
; CURRENT FILING DATE: 2003-07-16
; PRIOR APPLICATION NUMBER: 10/118,495
; PRIOR FILING DATE: 2002-04-08
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 32
; TYPE: DNA
; ORGANISM: Sinorhizobium meliloti
US-10-620-914-32

QY 806 TGTTCGCGCCACCTTCTGTACACGAGTCTGTGGGAGGAGCCCGGAGCGGATATGGAGG 865
Db 116 TGCTCTTTCCGAGCTCGTCTACCGCAGATCTGGGAGGAGCCGATTTGCGACATGGAAG 175
QY 866 TGATGAGATCAACCCCAAGGACACGGTGTGACCTGTAGCTAGCGGCGGTGCAATGCC 925
Db 176 CGATGAGATCCGCTCCCGGACATCGGATCGTACGATCGGTTCCGCGGCTGCAACATGC 235
QY 926 TGAACCTGTGTGAGGCGCGGCGGAGTGTGTGCTGCTGAGTGTGAGGAGTGTGGC 985
Db 236 TGACCTATCTCTCCGCGAGCTTCGCGGATAGACGTTGCGATCTCAACCCCATCA 295
QY 986 CGCGCTTCTGAGCTGGAAGAGTGGCCATTCAGACGCTGAGTGTGAGGAGTGTGGC 1045
Db 296 TCGCGCTCAACCGCTGAAAGTGTCTGCTTCCGCACTTCGCGGAGCCGACAGGAGCTGG 355
QY 1046 AGCTGTTCGCGAGGCGGTGCACCCCGCATTTAGGAGTGTGTACGA 1091
Db 356 TGGGTTCTCTCGCGTGAAGGTACGCGCAAGTGGCAAGTGGCAAGGCGCTTACCAAC 70

RESULT 11

US-10-118-495-22
; Sequence 22, Application US/10118495
; Publication No. US20030074688A1
; GENERAL INFORMATION:
; APPLICANT: Benning, Christoph
; APPLICANT: Riekhof, Wayne
; APPLICANT: Klug, Rouven
; TITLE OF INVENTION: Compositions and Methods for the Production of Betaine Lipids
; FILE REFERENCE: MSU-06897
; CURRENT APPLICATION NUMBER: US/10/118,495
; CURRENT FILING DATE: 2002-04-08
; PRIOR APPLICATION NUMBER: 60/283,812
; PRIOR FILING DATE: 2001-04-13
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 22
; LENGTH: 1248
; TYPE: DNA
; ORGANISM: Mesorhizobium loti
US-10-118-495-22

QY 704 TCCCTTACGTGCGGTGGCTGGCGGCCCCCTACTACGTGTGGATGGCGGCTGCCAGCG 763
Db 11 TCTCTCTCGGATCTGGTTTTCGCGCGGCAAGGAGTGGAAAGGCGGCTTACCAAC 70

```
QY 764 TTGGCCACGCGCTGCACGAGGAGCGCGTGGAGCGCGCCCATGTTCCCGCCCACTTCC 823
Db 71 GCGCGCTTTCAAAGCGGATCTCCGAGCGGTGTTGCGCTTCTGCTTTTCGGGCTCG 130
QY 824 TGTACACGAGTGTGTTGGAGGAGCCCGAGCGGATATGGAGGTGATGAGATCAACCCCA 883
Db 131 TCTATCGCAGATCTGGGAGAGCCCGATGTCACATGGAGGCGATGACGCTTGGTCAGG 190
QY 884 AGACACGGTGTGACCTGACTAGCGGCGGCTGCATGCCCTGAACTGCTGGTCAGG 943
Db 191 GCCATCGCATCGTCAACAATCGCTTCGCGCGGCTGCAACATCTCTGCGCTACCCGGTT 250
QY 944 GCGCGCGCAGGTGTTGCTGGTGGACTGCAACCCCGCGAGTGGCGCTTCTGGAGCTGA 1003
Db 251 CGCGGACGAGTGCAGCGCGTGCACCTCAACCGCGCCACATCGCGCTGAACCGCATGA 310
QY 1004 AGAAGTGGCCATTCAGCAGCTGGAGTTTGAAGAG---TGTGGCAGCTGTTCCGGAGG 1060
Db 311 AGCTGGAGCGGTGCGCGCTGCGCTTCGAGGCGGATCTGTTCCGCTTTTTCGGGCGCG 370
QY 1061 GCGTGACCCGCGATTTGAGGAGCTGTACGAGAGAGCTGGCGCGCTTCTGTCGCAAA 1120
Db 371 CCGACACAGCCCAATTCGCAAGCCATACCGCTTTATTGCGCGCGCATCTCGATCCGG 430
QY 1121 CCAGCCCAAACTTCTGCTCAAGCGCTCTGGTACTTCCAGCAGCGCTGTACTACCAAG 1180
Db 431 TCAGCCGCACTATTGGAGCGCGGCAACTGGCGTGTGTCGCGCGCATCGCGCTCTCG 490
QY 1181 CGCGCATGGCAAGCTGTGCTGGGTGCTGCAAGTCTGCGCGCTGGGTGCTGGGATCGGCA 1240
Db 491 ACCGCAATTTCTACGAGAC---GGCCTGCTCGCGCTGTTTATCGCCATGGCCATCGCA 547
QY 1241 AGACCGTCAAGCGCTCGCAACCGCCCAATGAGAGGAGCAGCGCGCTGTTGGGACA 1300
Db 548 CGCGCAAAATTTCTGGCGGTCAACCCCGGCCACATGATGGAAGCCAGGAATATCGCGAGC 607
QY 1301 GCAACATGCTCATCACTCTGTGAAGAAAGCGGCCAAGCGCTGTGTGGCTGTGCTCA 1360
Db 608 AGCGCGCTTCTTCAACGAG---AGCTGGCGCGGTCTTTCAGCAAGA 652
QY 1361 AGTTGAGCGCTGTGCTCTTCAACAAAGCCGCTGTGTGGTGGCGCGGCTGCGCG 1420
Db 653 AGCTTTTGAATGGCGACCTCGCGTAAAGCTGCTGCTGTTTCGGCTCGGCATTCGCGCG 712
QY 1421 GCAAGAGTACGCGCTGATCAAGCGGACGCGATCCCGCATGGAATACATCGCGGCA 1480
Db 713 CGCAGTACGATTCCTGATCACTCAGCGCAGCGCACCATGGCCAGCGTTCGAAAGCC 772
QY 1481 CCATGGACGCGTGGCGGAGAACTCGCACGCTGGCGAAGCAGAACTACTTCTACTACAA 1540
Db 773 GGCTGGAAGAGCTCGCTGCGATTTTCCCTGGAAACAAATTTTCCGCTGGCAGGCTT 832
QY 1541 GCCTCACCGGCAAGTTCCTGCGGACAACTGC-----CCCACTACTCGCGAGCGCG 1594
Db 833 TTGCGCGCGCTATCCAAATCCCGGTGAGCGCGCGCTGCGCGCTATCTGGAAGAGCA 892
QY 1595 CTTTCGCGCACCT-----CAAGAGTGGGTGGTGGCAACCTGACCGTCTCCACCACT 1648
Db 893 ACTACGAAACCATCCGCGGCAATATCGACGCGTTCGCGCATCCACCATGCAATCTGATCG 952
QY 1649 TCTTCATGGAGGAGCTCAAGCGCGCACCTACACCAAGGTGATTCGATGAGCACCGTGG 1708
Db 953 AATTCTCGCGGCAAGACGCGGCGACCGTGCATCGCTTATCTGCTGATGCGCAGG 1012
QY 1709 ACTGGCTGGATATGCCGTGGCCCAACGAGTGGCGCGGAGTGTGCTGGCCAGCAGTTGCG 1768
Db 1013 ACTGGATGACGATGACAGCTCAACGCGTGTGTTGCGAAATCAGCGCGACCGCTCG 1072
QY 1769 CGGCGGCGATGATCTGGGCGCTCGCGCTCC 1800
Db 1073 CAGCGCGCGGCTATCTTCCGCAACCGCGCGC 1104
```

```
RESULT 12
US-10-620-914-22
; Sequence 22, Application US/10620914
; Publication No. US20040093639A1
; GENERAL INFORMATION:
; APPLICANT: Benning, Christoph
; APPLICANT: Riekhof, Wayne
; APPLICANT: Klug, Rouven
; TITLE OF INVENTION: Compositions and Methods for the Production of Betaine Lipids
; FILE REFERENCE: MSU-07769
; CURRENT APPLICATION NUMBER: US/10/620,914
; CURRENT FILING DATE: 2003-07-16
; PRIOR APPLICATION NUMBER: 10/118,495
; PRIOR FILING DATE: 2002-04-08
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 22
; TYPE: DNA
; ORGANISM: Mesorhizobium loti
US-10-620-914-22
```

Query Match 3.6%; Score 69.6; DB 7; Length 1248;
Best Local Similarity 45.0%; Pred. No. 3.1e-08;
Matches 500; Conservative 0; Mismatches 579; Indels 33; Gaps 5;

```
QY 704 TCCCTACGTGCGCGTGGCTGCGCGCCCTTACTACGTGTGGATTTGGCGGCTGCGCCAGCG 763
Db 11 TCTCCTCGGATCTGGTTTTTTCGCGCGGCAAGGAAGTTGGAAGGCGCTTACAGAAC 70
QY 764 TTGGCCACGCGCTGCACGAGGAGCGGTGGAGCGCGCGCCCATGTTCCCGCCCACTTCC 823
Db 71 GCGCGCTTTCCAAAGCGGCGATCTCCGAGCGGTGTTTCGCTTCTGTTTTCGGGCTCG 130
QY 824 TGTACACGAGTGTGTTGGAGGAGCCCGAGCGGATATGGAGGTGATGAGATCAACCCCA 883
Db 131 TCTATCGCAGATCTGGGAGAGCCCGATGTCACATGGAGGCGCATGACGCTTGGTCAGG 190
QY 884 AGACACGGTGTGACCTGACTAGCGGCGGCTGCAATGCCCTGAACTGCTGCTGTCAGG 943
Db 191 GCCATCGCATCGTCAACAATCGCTTCGCGCGGCTGCAACATCTCGCTTACTTACCCGTT 250
QY 944 GCGCGCGCAGGTGTTGCTGGTGGAGTGTCAACCCCGCGCAGTGGCGCTTCTGGAGCTGA 1003
Db 251 CGCGGACGAGTGCAGCGCGTGCACCTCAACCGCGCCACATCGCGCTGAACCGCATGA 310
QY 1004 AGAAGTGGCCATTCAGCAGCTGGAGTTTGAAGAG---TGTGGCAGCTGTTTCGGGAGG 1060
Db 311 AGCTGGAGCGGTGCGCGCTGCGCTTCGAGGCGGATCTGTTCCGCTTTTTCGGCGCGCG 370
QY 1061 GCGTGACCCGCGATTTGAGGAGCTGTACGAGAGAGCTGGCGCGCTTCTGTCGCAAA 1120
Db 371 CCGACACAGCCCAATTCGCAAGCCATACCGCTTTATTGCGCGCGCATCTCGATCCGG 430
QY 1121 CCAGCCCAAACTTCTGCTCAAGCGCTCTGGTACTTCCAGCAGCGCTGTACTACCAAG 1180
Db 431 TCAGCCGCACTATTGGAGCGCGGCAACTGGCGTGTGTCGCGCGCATCGCGCTCTCG 490
QY 1181 CGCGCATGGCAAGCTGTGCTGGGTGCTGCAAGTCTGCGCGCTGGGTGCTGGGATCGGCA 1240
Db 491 ACCGCAATTTCTACGAGAC---GGCCTGCTCGCGCTGTTTATCGCCATGGCCATCGCA 547
QY 1241 AGACCGTCAAGCGCTCGCAACCGCCCAATGAGAGGAGCAGCGCGCTGTTGGGACA 1300
Db 548 CGCGCAAAATTTCTGGCGGTCAACCCCGGCCACATGATGGAAGCCAGGAATATCGCGAGC 607
QY 1301 GCAACATGCTCATCACTCTGTGAAGAAAGCGGCCAAGCGCTGTGTGGCTGTGCTCA 1360
Db 608 AGCGCGCTTCTTCAACGAG---AGCTGGCGCGGTCTTTCAGCAAGA 652
QY 1361 AGTTGAGCGCTGTGCTCTTCAACAAAGCCGCTGTGTGGTGGCGCGGCTGCGCG 1420
```

Db 653 AGCTTTTGAATGGGCGACCTCGGTAAGCGCTCGCTGTTCCGCGCTCGGCAATTCGCCGG 712
Qy 1421 GCAAGCAGTACGGGCTGATCAAGCGGAGCGATCCCTCAATGAGAACTACATCGCGCGCA 1480
Db 713 CGCAGTACGATTCCCTGATCACCTCAGGCGAGCGCAACCATGCGCAGGTTCTGAAGGCC 772
Qy 1481 CCATGACGCGTGGCGGAGAACTCGACGTGGCGAGCAGAACTACTTCTACTACAACT 1540
Db 773 GGCTGGAAAGAGCTCGCTCGGATTTTCCCTCGGAAACAAATTTTCCGCTGGCAGGCTT 832
Qy 1541 GCCTCACCGGCAAGTTCTCGCGGCAACTGC-----CCACCTACCTGCGCGAGCGG 1594
Db 833 TTGCGCGCGCTATCCAAATCCCGTGAGCGCGCCCTGCGCGCTATCTGAAAGAGAA 892
Qy 1595 CTTTCGCCACCTT-----CAAGAGTGGGTGGGACAACTGACCGTCTCCACCAACT 1648
Db 893 ACTACGAACCATCCGCGGCAATATCGACCGGTGCGCATCCACCATGCAATCTGATCG 952
Qy 1649 TCTTCATGAGGAGCTCAAGCGGACCTACACCAAGGTGATTCGTATGAGCACCGTGG 1708
Db 953 AATTCTCGCGGCAAGGACGCGGACCGTGCATCGCTTCACTCTGCTGATGCGCAGG 1012
Qy 1709 ACTGGCTGGATATGCCGTGGCCAAAGAGCTGCGCGAGTGCCTGGCCAAAGCAGTTGCGC 1768
Db 1013 ACTGGATACCGATGACGACTCAAGCGCTGTGGTGGAAATCAGCGGACCGCCTCG 1072
Qy 1769 CGGCGCGCATCGTATCTGCGCTCGCGCTCC 1800
Db 1073 CAGGCGCGCGTCACTCTTCGCAACCGCGCC 1104

RESULT 13

US-10-204-434A-12
; Sequence 12, Application US/10204434A
; Publication No. US20030170662A1
; GENERAL INFORMATION:
; APPLICANT: SETO, Haruo
; APPLICANT: KUZUYAMA, Tomohisa
; TITLE OF INVENTION: A Method for Screening A Substance Which Specifically Inhibits No
; FILE OF INVENTION: Mevalonate Pathway
; FILE REFERENCE: P22747
; CURRENT APPLICATION NUMBER: US/10/204,434A
; CURRENT FILING DATE: 2002-08-30
; PRIOR APPLICATION NUMBER: PCT/JP01/01501
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 1104
; TYPE: DNA
; ORGANISM: Streptomyces sp.
US-10-204-434A-12

Query Match 3.5%; Score 67.8; DB 6; Length 1104;
Best Local Similarity 45.1%; Pred. No. 9.3e-08;
Matches 252; Conservative 0; Mismatches 307; Indels 0; Gaps 0;
Qy 1344 GGTGTGGCTGTTCTCGTCAAGTTCTGAGCTGTGCTCTTCAAGCGCGTGTGTT 1403
Db 513 GGTCCGCGAGATCGAGAGATCGGCGGCGCTCGACATCCCGTGTATGTCGAGGAGT 572
Qy 1404 CGCGCGCGGCTCGCGGCAAGCAGTACGCTGTGATCAAGCGGAGCGGATCCCAATTGA 1463
Db 573 CGGCAACGGCTGAGCGGCGAGACCATCTGCTGCTCGCGACCTCGGCTGCGAGCGG 632
Qy 1464 GAATACATCGCGGACCACTGACGCGGTGGCGGAGAACTGCACGTGGCGAGCAGAA 1523
Db 633 GGACGTGAGCGGCGGCGGCGACGAGCTTCGCGCGCATCGAAGCGCGCGGAGCT 692
Qy 1524 CTACTTCTACTACAACTGCTCACCGGCAAGTTCTCTGCGCGCAACTGCGCCCACTTACCT 1583
Db 693 CGCGGATACGCGTTCTGTCAGCGGCTGGGGGAGTGTCCACGCGCGCTGCTGTGAGACG 752

Qy 1584 GCGCGAGCGCGCTTCGCCACCTCAAGAGTGGCGTGGGACAACTGACCGTCTCCAC 1643
Db 753 CCAGGACATCTCCCTCGCGCTCTCGCTCCGCGGTGCGTCAACCGCTCGACGTTGT 812
Qy 1644 CAACTTCTCATGAGGAGCTCAAGCGGCGCACCTACACCAAGGTGATTTCTGATGACCA 1703
Db 813 CGCGCGCTCGCGCTCGCGCGCGCGCTCGCTCTCTCGCGCGCTTCCTGCGCACCT 872
Qy 1704 CBTGAGTGGCTGGATATGCCGTGGCCCAACGAGCTGGCGAGTGGCTGGCCAAAGCAGT 1763
Db 873 GATGAGCAGCGGCTCGACGCGCTGATCAGAGCTCAGACCTGCTGACCGAGTGGC 932
Qy 1764 TGGCGCGGCGGAGTCTGATCTGCGCTCGCTCCCTCAGCCCGCTTACGCGAGCT 1823
Db 933 GCGCTGTCAGACCATCTCGCGCGCGCACCCCGCGCGACCTCACCGCTGCGACGCT 992
Qy 1824 GATCCAGAGGCGGCTTCGAGCTGCGTGCATCCGCGCGCCACTCAGGGCTACATGA 1883
Db 993 GCTCCACGCGGAGCTCGTGACTTTCGCGCGACCGGGGATCGACACGCGCGCTCGC 1052
Qy 1884 CCGCGTCAACATGTACAGC 1902
Db 1053 CCAGCGCTCCAGTCCATC 1071

RESULT 14

US-09-918-740-57
; Sequence 57, Application US/09918740
; Publication No. US20030033626A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Frederick
; APPLICANT: Kuehnle, Adelheid
; TITLE OF INVENTION: Manipulation of genes of the mevalonate and isoprenoid pathways t
; FILE OF INVENTION: create novel traits in transgenic organisms
; FILE REFERENCE: KAS-103XC1
; CURRENT APPLICATION NUMBER: US/09/918,740
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/221,703
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 76
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 57
; LENGTH: 6798
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Streptomyces sp CL190 gene cluster containing mevalonate pathway
; OTHER INFORMATION: and IPP isomerase orfs
US-09-918-740-57

Query Match 3.5%; Score 67.8; DB 3; Length 6798;
Best Local Similarity 45.1%; Pred. No. 1.4e-07;
Matches 252; Conservative 0; Mismatches 307; Indels 0; Gaps 0;
Qy 1344 GGTGTGGCTGTTCTCGTCAAGTTCTGAGCTGTGCTCTTCAAGCGCGTGTGTT 1403
Db 3835 GGTCCGCGAGATCGAGAGATCGCGCGCGCTCGACATCCCGTGTATGTCGAGGAGT 3894
Qy 1404 CGCGCGCGGCTCGCGGCAAGCAGTACGCTGTGATCAAGCGGAGCGGATCCCAATTGA 1463
Db 3895 CGGCAACGGCTGAGCGGCGAGACCATCTGCTGCTGCGCGACCTCGGCTGAGCGGCG 3954
Qy 1464 GAATACATCGCGCGCACCATGAGCGGCTGGCGAGAACTGCACGTGGCGAGCAGAA 1523
Db 3955 GGACGTGAGCGCGCGCGGCGGCGAGCTTCGCGCATCGAAGCGCGCGGAGCT 4014
Qy 1524 CTACTTCTACTACAACTGCTCACCGGCAAGTTCTGCGCGCAACTGCGCCCACTTACCT 1583
Db 4015 CGCGGACTACGCGTTCTGTCAGCGTGGGGGAGTCCACCGCGCTGCTGTGAGCGC 4074
Qy 1584 GCGCGAGCGCGCTTCGCCACCTCAAGAGTGGCGTGGGACAACTGACCGTCTCCAC 1643
Db 4075 CCAGGACATCTCTCTGCGCGCTCTCGCTCCGCGCGTGTGCGTCAACCGCTCGACGTTGT 4134

QY 1644 CAACTTCTTCATGAGGAGCTCAAAGCGCGCACTACACCAAGGTGATTCTGATGACCA 1703
Db 4135 CCGCGCCCTCGCGTCCGCGCCCGCGCTCGGCTCTCCGCGGCTTCTGCGCACCT 4194
QY 1704 CGTGGACTGGCTGATATGCGCTGGCGCAACGAGCTGGCGAGTGGCTGGCCCAAGCAGGT 1763
Db 4195 GATGGACGACGGCGTGCACGGCTGATCAGAGCTCACAGCTGGCTGGACCAAGCTGGC 4254
QY 1764 TGGCGCGGGCGGATGCTGATCTGGCGCTCCGCTCCCTCAGCCGCGCTTACGCCGAGCT 1823
Db 4255 GCGCTGACAGACCATGCTCGCGCGCGCACCCGCGCGACCTCACCGCTGCGACGCTGCT 4314
QY 1824 GATCCAGAAGCGGGCTTCAGCTGGCTGCATCCGCGCGGCACTCAGGGCTTACATGGA 1883
Db 4315 GCTCCACGGCGAGCTGCGTGACTTCTGCGCGGATCGACACGCGCGCTCGC 4374
QY 1884 CCGCGTCAACATGTACAGC 1902
Db 4375 CCAGCGCTCCAGCTCCATC 4393

RESULT 15

US-10-204-434A-13
; Sequence 13, Application US/10204434A
; Publication No. US20030170662A1
; GENERAL INFORMATION:
; APPLICANT: SETO, Haruo
; APPLICANT: KUZUYAMA, Tomohisa
; TITLE OF INVENTION: A Method For Screening A Substance Which Specifically Inhibits Nc
; TITLE OF INVENTION: Mevalonate Pathway
; FILE REFERENCE: P22747
; CURRENT APPLICATION NUMBER: US/10/204,434A
; CURRENT FILING DATE: 2002-08-30
; PRIOR APPLICATION NUMBER: PCT/JP01/01501
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13
; LENGTH: 6798
; TYPE: DNA
; ORGANISM: Streptomyces sp.
US-10-204-434A-13

Query Match 3.5%, Score 67.8; DB 6; Length 6798;
Best Local Similarity 45.1%; Pred. No. 1.4e-07;
Matches 252; Conservative 0; Mismatches 307; Indels 0; Gaps 0;
QY 1344 GGTGTGCTGTTCGTCAAGTTTCGTGAGCTGGTCTCTTCAACAAGGCGCTGTGTGTT 1403
Db 3835 GGTCCCGAGATCAGAGATCGCGCGGCGCTCGACATCCCGTGATCGTCAAGGAGT 3894
QY 1404 CGCGCGCGGTGCGCGCAAGCAGTACGCGCTGATCAAGCGGACCGCATCCCATTTGA 1463
Db 3895 CGGCAACGGCTGAGCGCGCAGACCATCTGCTCTCGCGACCTCGCGCTGCGAGCGC 3954
QY 1464 GAATACATCGCGCGCACCATGAGCGGTGGCGGAACTCGCACCTGGCGCAGAGAA 1523
Db 3955 GGACGTGAGCGCGCGCGCGCACGGAATTCGCGCGCATCGAGAACGCGCGCGGAGCT 4014
QY 1524 CTACTTCTACTACAACTGCCTCACCGCAAGTTCCTCGCGGACAACTGCGCCCACTACCT 1583
Db 4015 CGGGCATCTACCGTTCTTGACGGCTGGGGGAGTCCACCGCGCTTGCCTGTGGAGCG 4074
QY 1584 GCGCGAGCGGCGCTTCGCCACCTCAAGAGTGGGTGGTGGACAACTGACCGCTCTCCAC 1643
Db 4075 CCAGGACATCTCCCTCGCGCTCTCGCTCGCGGGGTGGGTCAACCGCTCGACGCTGGT 4134
QY 1644 CAACTTCTTCATGAGGAGCTCAAAGCGCGGACCTACACCAAGGTGATTCTGATGACCA 1703
Db 4135 CCGCGCGCTTCGCGTCCGCGCGCGCGCTCGGCTCTCTCGCGCGCTTCTCTGCGCACCT 4194
QY 1704 CGTGGACTGGCTGGATATATCCCGTGGCGGCAACGAGCTGGCGGAGTGCCTGGCCCAAGCAGGT 1763

Db 4195 GATGGACGACGGCGTGCACGGCTGATCAGAGCTCACAGACCTGGCTGGACCAAGCTGGC 4254
QY 1764 TGGCGCGGGCGGATGCTGATCTGGCGCTCCGCTCCCTCAGCCGCGCTTACGCCGAGCT 1823
Db 4255 GCGCTGACAGACCATGCTCGCGCGCGCACCCGCGCGACCTCACCGCTGCGACGCTGCT 4314
QY 1824 GATCCAGAAGCGGGCTTCAGCTGGCTGCATCCGCGCGGCACTCAGGGCTTACATGGA 1883
Db 4315 GCTCCACGGCGAGCTGCGTGACTTCTGCGCGGATCGACACGCGCGCTCGC 4374
QY 1884 CCGCGTCAACATGTACAGC 1902
Db 4375 CCAGCGCTCCAGCTCCATC 4393

Search completed: March 14, 2006, 04:22:33
Job time : 1630 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 14, 2006, 01:11:21 ; Search time 659 Seconds
(without alignments)
6816.854 Million cell updates/sec

Title: US-10-620-914-44

Perfect score: 1947

Sequence: 1 atgggggtcgggtcgtgacgg.....gcgccaaagaagacaactaa 1947

Scoring table: IDENTITY NUC

Gapop 10_0 , Gapext 1.0

Searched: 7673375 seqs, 1153648444 residues

Total number of hits satisfying chosen parameters: 15346750

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

- Database : Published Applications NA_New.*
- 1: /cgn2_6/prodata/1/pubpna/US06_NEW_PUB.seq.*
 - 2: /cgn2_6/prodata/1/pubpna/US06_NEW_PUB.seq.*
 - 3: /cgn2_6/prodata/1/pubpna/US07_NEW_PUB.seq.*
 - 4: /cgn2_6/prodata/1/pubpna/PCT_NEW_PUB.seq.*
 - 5: /cgn2_6/prodata/1/pubpna/US09_NEW_PUB.seq.*
 - 6: /cgn2_6/prodata/1/pubpna/US09_NEW_PUB.seq.*
 - 7: /cgn2_6/prodata/1/pubpna/US10_NEW_PUB.seq.*
 - 8: /cgn2_6/prodata/1/pubpna/US10_NEW_PUB.seq.*
 - 9: /cgn2_6/prodata/1/pubpna/US11_NEW_PUB.seq.*
 - 10: /cgn2_6/prodata/1/pubpna/US11_NEW_PUB.seq.*
 - 11: /cgn2_6/prodata/1/pubpna/US11_NEW_PUB.seq.*
 - 12: /cgn2_6/prodata/1/pubpna/US11_NEW_PUB.seq.*
 - 13: /cgn2_6/prodata/1/pubpna/US60_NEW_PUB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	66	3.4	1577	9	US-11-096-568A-10906
2	64	3.3	2736	8	US-10-858-730-38
3	60.8	3.1	1325	9	US-11-096-568A-1077
4	59.6	3.1	1347	9	US-11-096-568A-12531
5	57.8	3.0	1599	9	US-11-096-568A-26187
6	57.6	3.0	1444	9	US-11-096-568A-20988
7	57.2	2.9	1806	12	US-11-166-609-1
8	56.2	2.9	3513	8	US-10-858-730-142
9	56	2.9	4282	12	US-11-183-624-1
10	56	2.9	4282	12	US-11-183-624-3
11	55.2	2.8	1243	9	US-11-096-568A-25531
12	55.2	2.8	88421	12	US-11-205-109-1
13	54.8	2.8	88421	12	US-11-205-109-1
14	54.6	2.8	1283	9	US-11-096-568A-23551
15	54	2.8	2025	12	US-11-167-048-2
16	53.8	2.8	2403	7	US-10-649-457-4
17	53.8	2.8	37507	8	US-10-522-037-2
18	53.6	2.8	1182	9	US-11-096-568A-26293
19	53.6	2.8	1580	9	US-11-096-568A-20206
20	53.6	2.8	2873	12	US-11-136-527-2776

21	53	2.7	1292	9	US-11-096-568A-19003	Sequence 19003, A
22	52.8	2.7	1958	9	US-11-096-568A-26982	Sequence 26982, A
23	52.8	2.7	3387	12	US-11-091-883-183	Sequence 183, App
24	52.8	2.7	3439	12	US-11-000-688-851	Sequence 851, App
25	52.6	2.7	1572	9	US-11-096-568A-19140	Sequence 19140, A
26	52.2	2.7	1859	9	US-11-096-568A-25607	Sequence 25607, A
c	27	30191	7	US-10-330-773-631	Sequence 631, App	
28	52	2.7	1401	9	US-11-096-568A-22550	Sequence 22550, A
29	52	2.7	3624	7	US-10-755-092-6	Sequence 6, Appli
30	51.2	2.6	1321	9	US-11-096-568A-20876	Sequence 20876, A
31	51.2	2.6	1566	12	US-11-094-917-55	Sequence 55, Appli
32	51.2	2.6	1816	12	US-11-094-917-8	Sequence 8, Appli
33	51.2	2.6	6177	12	US-11-094-917-7	Sequence 7, Appli
34	50.8	2.6	906	12	US-11-108-163B-2	Sequence 2, Appli
35	50.8	2.6	906	12	US-11-108-163B-4	Sequence 4, Appli
36	50.8	2.6	1335	8	US-10-750-185-48926	Sequence 48926, A
37	50.8	2.6	1335	8	US-10-750-623-48926	Sequence 48926, A
38	50.8	2.6	1375	12	US-11-108-163B-1	Sequence 1, Appli
39	50.6	2.6	1866	9	US-11-096-568A-13310	Sequence 13310, A
40	50.4	2.6	1530	7	US-10-501-035-58	Sequence 58, Appl
41	50.4	2.6	1588	9	US-11-096-568A-23028	Sequence 23028, A
42	50.4	2.6	1630	9	US-11-096-568A-17222	Sequence 17222, A
43	50.4	2.6	1818	9	US-11-096-568A-21800	Sequence 21800, A
44	50.4	2.6	2487	9	US-11-096-568A-14592	Sequence 14592, A
45	50.4	2.6	2506	9	US-11-096-568A-19863	Sequence 19863, A

ALIGNMENTS

RESULT 1

US-11-096-568A-10906
; Publication 10906, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE OF INVENTION: Therby
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 10906
; LENGTH: 1577
; TYPE: DNA
; ORGANISM: Triticum aestivum
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1577)
; OTHER INFORMATION: Ceres Seq. ID no. 13597454
US-11-096-568A-10906

Query Match 3.4%; Score 66; DB 9; Length 1577;
Best Local Similarity 47.2%; Pred. No. 2.5e-06;
Matches 201; Conservative 0; Mismatches 225; Indels 0; Gaps 0;

QY	1408	GGCGGCGTCCCGGCAAGCAGTACGCGCTGATCAAGGGCGGCGCATCCCATTTGAGAAC	1467
DB	315	GGATACATGCGCGGCGGCTCTACGACCTGGACGGCTCCAGTACGGCAGCGGGTGGAG	374
QY	1468	TACATGCGCGCACCATGAGCGCGTGGCGGAGAACTCGCACTGCGCAAGCAGAACTAC	1527
DB	375	CTCAGGTGCTCATCGCGGCTTCGGTGCAGAGGCGCATCAAGTGCCTGCCGACATCGTC	434
QY	1528	TTCTACTACACTGCTCATCCGCGAAGTTCCTGCGGGGACAACTGCCCGACCTACCTCGGC	1587
DB	435	ATCAACCAACCGTGGCGCGGAGCAAGGACGGCGCGGCTCTACTGCACTCTTCAAGGGC	494
QY	1588	GAGCGCGCTTCGCGCACCTCAAGAGTGGCGTGGTGACAACTGACCTCTCCACCAAC	1647
DB	495	GGAGGACCGAGGGGCTCTCTGACTGGGTGCGGGCATGATCTGACGACGACCAAC	554
QY	1648	TTCTTTCATGAGGAGCTCAAGCGCGCACTACACCAAGGTGATTTCTGATGGACCACTG	1707

Db 555 TTCTCCGACGCGCGGTACCGTGATACCGCGCGGACCTTCGCGCGCGCGGACATC 614
Qy 1708 GACTGGCTGATATGCCCGTGGCCACAGAGCTGGCGAGTGCTGGCCGAAGAGGTGGG 1767
Db 615 GACCACCTCAACCGCGCGTCCAGAAAGAGAGCTCTCCGACTGGCTCAAGTCC 674
Qy 1768 CCGGGGGCATGTCATCTGGCGCTCGCGCTCCCTCAGCCGCGCTACGCGGAGCTGATC 1827
Db 675 GACCTCGCTTCAGCGCTGGCGCTCGACTTCGCCAAGGGCTACTCGGCCGAGCTGGCC 734
Qy 1828 CAGAAG 1833
Db 735 AGGACG 740

RESULT 2

US-10-858-730-38
; Sequence 38, Application US/10858730
; Publication No. US20050255568A1
; GENERAL INFORMATION:
; APPLICANT: Bailey, Richard B.
; APPLICANT: Blomquist, Paul
; APPLICANT: Doten, Reed
; APPLICANT: Driggers, Edward M.
; APPLICANT: Madden, Kevin T.
; APPLICANT: O'Leary, Jessica
; APPLICANT: O'Toole, George
; APPLICANT: Trueheart, Joshua
; APPLICANT: Walbridge, Michael J.
; APPLICANT: Yorgey, Peter S.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR AMINO ACID
; TITLE OF INVENTION: PRODUCTION

FILE REFERENCE: 14184-030001
CURRENT APPLICATION NUMBER: US/10/858,730
CURRENT FILING DATE: 2004-06-01
PRIOR APPLICATION NUMBER: US 60/475,000
PRIOR FILING DATE: 2003-05-30
PRIOR APPLICATION NUMBER: US 60/551,860
PRIOR FILING DATE: 2004-03-10
NUMBER OF SEQ ID NOS: 364
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 38

LENGTH: 2736

TYPE: DNA

ORGANISM: Streptomyces coelicolor

US-10-858-730-38

Query Match 3.3%; Score 64; DB 8; Length 2736;
Best Local Similarity 47.6%; Pred. No. 7.6e-06;
Matches 221; Conservative 0; Mismatches 240; Indels 3; Gaps 1;
Qy 1458 CATTGAGAACTACATCGCGCGCACCATGGAAGGGCTGGCGGAGAACTCGCAGTGGCCAA 1517
Db 1011 CACCAAGCAGCGCTCCCAAGGCGCACCCGCCAGAGGAGCGCGGACTACCTCGGCAC 1070
Qy 1518 GCAGAACTACTTCTACTACAACTCCCTCAGCGCAAGTCTTGGCGGACAACTGCCCCAC 1577
Db 1071 CGCCAGCTCATGACAGCTGGCGCATCTCGTCCAGACCTCGCTGGCGGAAACACCGCGCGG 1130
Qy 1578 CTACTCGCGGAGCGCGCTTCGCCACCTTCAAGAGTGGGCTGTGG---ACAACCTGAC 1634
Db 1131 CTGTTCCGCGAGCGGCGCTTCGCCCGGCTCGCCGACCATCCGACCTTGGCGGCTTCCCA 1190
Qy 1635 CGTCTCCACCAACTTCTTTCATGAGAGAGCTCAAAGCGCGCACTTACACCAAGGTGATCT 1694
Db 1191 GCTCGCCACCATGAGCTCGCGGAGCAGCGCGGAGCCACCAACCGCCCTCGGCCAGCT 1250
Qy 1695 GATGACCACTGACTGGCTGGATATGCCCGTGGCCCAAGAGCTGGCGGAGTGCCTGGC 1754
Db 1251 CTTGACCGGCTCGCGGAGAGTCTTGGCGTACGCCGATGCGCGGAGTACCGCAC 1310
Qy 1755 CAAGCAGTTTGGCGCGCGCGCATCGTCACTTGGCGCTCGGCTCCCTCAGCCGCGCCTA 1814

Db 1311 CAAAGTCTCTCGCCCAAGAACTGGCTCCCGCAGCGCGTGGCCCCCAGCCCCCGCT 1370
Qy 1815 CGCGAGCTGATTCAGAAAGCGGGCTTCAGCTGGCTGCATCGCGGCGGCGCACTCAGGG 1874
Db 1371 CGACGCGCGCGGAGAGACCTCGGCGCTTTCAGAGCCGTTCGCGCGCGCTGGAGGT 1430
Qy 1875 CTACATGGACCGGCTCAACATGTACAGCTCTCTTCTACATGGCCC 1918
Db 1431 CTTGCGCGCGGAGTTCATCGAGTCTTACATCATCTCCATGTGCC 1474

RESULT 3

US-11-096-568A-1077
; Sequence 1077, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; TITLE OF INVENTION: Therby
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 1077
; LENGTH: 1325
; TYPE: DNA
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1325)
; OTHER INFORMATION: Ceres Seq. ID no. 13599631
US-11-096-568A-1077

Query Match 3.1%; Score 60.8; DB 9; Length 1325;

Best Local Similarity 45.4%; Pred. No. 4.2e-05;

Matches 258; Conservative 0; Mismatches 307; Indels 3; Gaps 1;

Qy 1218 GCGCGTGTGTCTGGGACTGGGCAAGACCGTCAAGCGCTCGCCAAACGCGCCCAATGGA 1277
Db 293 GCGGTGGGTTCGGCAAGCCCGAGCGCGGAGTGCCTGCGCGGATCACTGCGCCACATCGG 352
Qy 1278 GAGCAGCGCGCTCTGTGGGACAGCAACATGCTCATCTCTGTGAAGAACCGGCCAA 1337
Db 353 CTTCCACCGCGCGACCTGCTGTCAGCTGCTCATCAAGAACCCCAACCGGTGCCCAT 412
Qy 1338 GCGCTGT 1397
Db 413 CCGCTGCTGACATCAGCTACCTCATCGACAGCGCGGAGGAGCTCTGCGGGGCT 472
Qy 1398 GTGGTTGGCGCGCGGTGCGCGGCAAGCAGTACGCGCTGA---TCAAGCGGACGCGCAT 1454
Db 473 CATCCGGAACCGCGGACCATTCGCGGCAAGCGGAGGAGACGCTCAAGTGGCCATCAC 532
Qy 1455 CCCATTGAGAACTACATCGCGGCAACCATGGAAGCGGTGGCGGAGAACTGCGACGTGGG 1514
Db 533 GCTCGACTTCGACGACATCGGAGCACCTACGCGGACATCAAGCGCGGAGCATATCCC 592
Qy 1515 CAAGCAGAACTACTTCTTCTAATACTGCTCAGCGCAAGTCTTCTGCGGCAACTGCC 1574
Db 593 CTACCTGCTCGGCTCATCTTCTCTGCGGAGCTGCGCGCTTTCGCGGCGCATCAAGATCCC 652
Qy 1575 CACTACTCGCGGAGCGGCTTCGCCACCTTCAAGAGTGGGTGTGGCAACACTGAC 1634
Db 653 GCTGGAAGTTCGGGAGATCCCATCCGCTACAGCCGACGCTGAGCGTGGACAAGAT 712
Qy 1635 CGTCTCCACCAACTTCTTCTTATGAGAGAGCTCAAAGCGCGCACTTACACCAAGGTGATCT 1694
Db 713 CAAAGTCCACCACTTCTTCTTCTGAGGAGACACGCGCCACCATCCACCTGAGCTGGAGAA 772
Qy 1695 GATGACCACTGAGCTGGCTGGATATGCCCGTGGCCAAAGAGCTGGCGGAGTGCCTGGC 1754
Db 773 CAAGAAGCAGTTTGACCTCGGCTCAACCTGCTCCAGTACGAGATGTGGCTCGCGGACGA 832

; TYPE: DNA
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1444)
; OTHER INFORMATION: Ceres Seq. ID no. 12391561
US-11-096-568A-20988

Query Match 3.0%; Score 57.6; DB 9; Length 1444;
Best Local Similarity 46.2%; Pred. No. 0.00024; Indels 0; Gaps 0;
Matches 192; Conservative 0; Mismatches 224; Indels 0; Gaps 0;

QY 1527 CTTCTACTACAACTGCTCACCAGGAGTTCTCGCGGCAAACTGCCCCACCTACTCTGCG 1586
DB 510 CTTCTGGCCCTACTGCTCAGCTTCTGCATGTGTATATCTTCCCGGCACTCTCTGCG 569
QY 1587 CGAGGGCGGCTTCCGCCACCTCAAGAGTGGGTGGCAAACTGACCTCTCCACCAA 1646
DB 570 CGACCCGCGCTTACCAGCGGCGACGGCAACACCTTCTATCTTCATGGAACAAGGACAA 629
QY 1647 CTTCTTCATGGAGAGCTCAAGCGCGCACCTACACCAAGGTGATTTCTGATGGACCACT 1706
DB 630 GGAATTTCTGCTCTCTCCAGCGGGCCCTCCACATCAACGCCCACTTCTATCGGCAACCA 689
QY 1707 GGAATGCTGATATGCTGCGGCGGCGAGTGGCGGAGTGGCTGCGGCAAGCAGGTTGC 1766
DB 690 CAACCCGGCCCTCAACCGGACTTCACTGGGTGCAGGCGCTGGGGGTCACTTCTGTCGC 749
QY 1767 GCGGGGGGCGATGCTCATATCTGGGGCTCCGCTCCCTCAGCCCGGCTTACCGGAGCTGAT 1826
DB 750 ACCAGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGAGTG 809
QY 1827 CCAGAGGGCGGCTTCAAGCTGGCTGCTATCCGCGCGGCGGCGGCGGCGGCGGCGGCGG 1886
DB 810 GAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 869
QY 1887 CGTCAACATGTACAGCTCTCTTACATGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1942
DB 870 CGTCAGAGCGGCGCTGGGCGTCCAGGCGGCTCCCGGGGCTCTCCGTCAGCGCA 925

RESULT 7
US-11-166-609-1
; Sequence 1, Application US/11166609
; Publication No. US20060015968A1
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, MARC C.
; APPLICANT: FOX, TIM
; APPLICANT: HUFFMAN, GARY
; APPLICANT: TRIMNELL, MARY
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES MEDIATING MALE FERTILITY AND
; TITLE OF INVENTION: METHOD OF USING SAME
; FILE REFERENCE: 1148CR
; CURRENT APPLICATION NUMBER: US/11/166,609
; CURRENT FILING DATE: 2005-06-24
; PRIOR APPLICATION NUMBER: 10/412,000
; PRIOR FILING DATE: 2003-04-11
; PRIOR APPLICATION NUMBER: 09/670,153
; PRIOR FILING DATE: 2000-09-26
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 1
; LENGTH: 1906
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1638)
US-11-166-609-1

Query Match 2.9%; Score 57.2; DB 12; Length 1906;
Best Local Similarity 43.1%; Pred. No. 0.00031;
Matches 275; Conservative 0; Mismatches 363; Indels 0; Gaps 0;

QY 908 GCGGCGGCTGCAATGCTTGAACCTTCTGCTGAGGGGGCGGCCAGGTGCTGCGTGG 967
DB 911 GCGGCGGCTTGGGGACGATAGAGCTTCCGGGACGCTGGTGTCAACTTCTGATCGCG 970
QY 968 ACTGCAACCCCGCGAGTGGCGCTTCTTGAGCTGAAGAAGTGGCCATTACAGAGCTGG 1027
DB 971 GCGGGACACGACGCGGCGAGCGCTGCTGCTGCTTACGACATGCGCATGTCACCCCG 1030
QY 1028 AGTTTGAAGACGTTGGCAGCTGTTCCGGCGAGGCGCTGACCCCGCATTTGAGAGCTGT 1087
DB 1031 ACCTGCGCGAGAGCTGCGCGCGGAGCTGTGCGCGGTTTCGAGGCGGAGCGCGCGCGAG 1090
QY 1088 ACCAGAGAAGCTTGGCGGCGCTTCTCTGTCGCAACACGACCACTTCTGCTTCAAGCGCC 1147
DB 1091 AGGCGCTCAGCTCTGCTCTGCGGCGGCGCTGACGCCGACGACAGGCGCTTCGCGGCC 1150
QY 1148 TCTGGTACTTCCAGCAGCGCTCTATACAGGGCGGCTATGGGCAAGCTGTGCTGGGTGC 1207
DB 1151 GCGTGGCGAGTTTCGCGGCGCTCTCACCTACGACAGCTTCGCGCAAGCTGTCTACCTCC 1210
QY 1208 TGCAAGTGGCGGCTGCTGGGAGTGGGCAAGACCGTCAAGCGCTTCGCAACCGCGC 1267
DB 1211 ACCTGCTGCTACCGAGACGCTCCGCTGTACCCCGCGCTTCCCTCAGGACCCCAAGGGA 1270
QY 1268 CCACAATGGAGGAGCAGCGCGCTCTGTGGGACAGCAACATGCTCTCATCTCTGTAAGA 1327
DB 1271 TCCTGGAGGACGAGCTGCTGCGGACGGAAGGTGAGGCGCGGCGGATGCTGACGT 1330
QY 1328 ACGGGCGGCAAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1387
DB 1331 ACCTGCGCTACTCGATGGGCGGATGAGTACAACTGGGGCGGCGGCGGCGGAGCTTCC 1390
QY 1388 AGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1447
DB 1391 GCGCGGAGCGGTGATCAACAGGATGCGCGCTTCGCAACGCGTTCGCTTCAAGTTCA 1450
QY 1448 ACGGCATCCCCATTGAGAACTACATCGCGCGCACCATGAGCGGCTGGCGGAGAACTCGC 1507
DB 1451 CGCGTTCCAGGCGGCGCGGAGGATCTGCTGGGCAAGGACTCGCGGTACTCGCATGA 1510
QY 1508 ACCTGCGCAAGCAGAACTACTTCTTACTACAACTGCGCTC 1545
DB 1511 AGATGGCGCTGGCATCTCTTCCGCTTCTACAGCTTC 1548

RESULT 8
US-10-858-730-142
; Sequence 142, Application US/10858730
; Publication No. US20050255568A1
; GENERAL INFORMATION:
; APPLICANT: Bailey, Richard B.
; APPLICANT: Blomquist, Paul
; APPLICANT: Doten, Reed
; APPLICANT: Driggers, Edward M.
; APPLICANT: Madden, Kevin T.
; APPLICANT: O'Leary, Jessica
; APPLICANT: O'Toole, George
; APPLICANT: Trueheart, Joshua
; APPLICANT: Walbridge, Michael J.
; APPLICANT: Yorgey, Peter S.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR AMINO ACID
; TITLE OF INVENTION: PRODUCTION
; FILE REFERENCE: 14184-030001
; CURRENT APPLICATION NUMBER: US/10/858,730
; CURRENT FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/475,000
; PRIOR FILING DATE: 2003-05-30
; PRIOR APPLICATION NUMBER: US 60/551,860
; PRIOR FILING DATE: 2004-03-10
; NUMBER OF SEQ ID NOS: 364
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 142

;	LENGTH: 3513	
;	TYPE: DNA	
;	ORGANISM: Streptomyces coelicolor	
US-10-858-730-142		
Query Match	2.9%;	Score 56.2; DB 8; Length 3513;
Best Local Similarity	44.6%;	Pred. No. 0.00055;
Matches 266;	Conservative 0;	Mismatches 328; Indels 3; Gaps 1;
QY	1146	CCTCTGTTACTTCAGACACGGCTGTACTACCAAGGCGGATCGGCAAGCTGTCTGGGT 1205
Db	1179	CCTCTGGTCGAATTACGTCGGCGGGACGGCTGCGCGACATGAGGAACTGGCCGCGC 1238
QY	1206	GCTGCAGTGCTGGCGGTGTCTGGGACTGGGCAAGACCGTCAAGCGCTCGCCACGC 1265
Db	1239	GTTCCGCCACCGCTCCACGCTGCGGATGCTCTGACTCCACCGAGTGCAGTCATCCG 1298
QY	1266	GCCCAAAATGGAGGAGCAGCGCGCTCTGTGGGACAGCAAATGCTCATCCACATTCGTGAA 1325
Db	1299	GGCGGGCTTGAGAAAGCTCGCGCGCCGCGGTGATCAACTCGGTCAACTACGAGGACGG 1358
QY	1326	GAACGGGCCCAAGCGCTGTGTGTGGTGTTCGTCAAGTTCGTGAGCGTGTGTCTTTCAA 1385
Db	1359	CGCGGGCCCCGAGTCCCGGTTCCGCCCGCTCAGCAAGCTCGCCGGGAGCACGGCGCGC 1418
QY	1386	CAAGGCGTCTGTGGTTTCGGCGGCGCGTGC CGGGCAAGCAGTACGCGCTGATCAAGC 1445
Db	1419	GCTGATCGCGCTGACCATCAGCAGGTGGACAGGCGCCGACCGCCGAGAAAGGTGCA 1478
QY	1446	GGACGGGCATCCCATTTGAGAACTACATCGCGCGCACCATGACGGGTGGCGGAGAACTC 1505
Db	1479	GATCGCGGAAGGCTCATCGACGACCTCAACCGGCACTGGGGCATCCACAGTCCGACAT 1538
QY	1506	GCAGTGGCGCAAGAGAACTACTTTACTATCAACTGTCTCACCGCAAGTTCTCGCGCGA 1565
Db	1539	CCTCGTCGACTGCTGACCTTACCTTCCACCATCTGCACCGGCC---AGGAGAGTCCCGCAAGGA 1595
QY	1566	CAACTGCCCACTACTCTGGCGGCGGCGCTTCGCCACCTCAAGAGTGGCTGTGTGGA 1625
Db	1596	CGCGCTTGGCCACCATCGAGGGCATTCGGGAACTCAAGCGGCGCAACCGGACGTGCAGAC 1655
QY	1626	CAACCTTGACCGTCTTCACCAACTTTCTTCATGAGGAGAGCTCAAAAGCGCGCACTTACACCAA 1685
Db	1656	CACGCTCGGCTGTGNAACATCTCTTCGGGCTCAACCGGCGCCCGCATCTGTCTCAA 1715
QY	1686	GGTGATTCTGATGGACCACTGGACTGTGTGGANATGCCCGTGGCCAAAGAGCTGGC 1742
Db	1716	CTCCGTCTTCTCGAGGAATTCGCTCAAGGCGCGCTGGACTCGGCGCATCTGTGACGCG 1772

LOCATION: (2077)..(3078)
OTHER INFORMATION: ORF 1; positive strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (3118)..(4032)
OTHER INFORMATION: ORF 2; positive strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (4038)..(5048)
OTHER INFORMATION: ORF 3; positive strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (6665)..(5814)
OTHER INFORMATION: ORF 4; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (7703)..(6693)
OTHER INFORMATION: ORF 5; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (9464)..(8130)
OTHER INFORMATION: ORF 6; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (9691)..(10761)
OTHER INFORMATION: ORF 7; positive strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (12751)..(10829)
OTHER INFORMATION: ORF 8; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (13617)..(12802)
OTHER INFORMATION: ORF 9; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (15203)..(13614)
OTHER INFORMATION: ORF 10; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (15591)..(15963)
OTHER INFORMATION: ORF 11; positive strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (15880)..(19035)
OTHER INFORMATION: ORF 12; positive strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (19032)..(39713)
OTHER INFORMATION: ORF 13; positive strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (39713)..(65800)
OTHER INFORMATION: ORF 14; positive strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (65826)..(66530)
OTHER INFORMATION: ORF 15; positive strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (66546)..(67370)
OTHER INFORMATION: ORF 16; positive strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (67384)..(70059)
OTHER INFORMATION: ORF 17; positive strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (70099)..(70662)
OTHER INFORMATION: ORF 18; positive strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (70659)..(71906)

OTHER INFORMATION: ORF 19; positive strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (73439)..(71964)
OTHER INFORMATION: ORF 20; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (74216)..(73563)
OTHER INFORMATION: ORF 21; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (75424)..(74213)
OTHER INFORMATION: ORF 22; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (75535)..(76464)
OTHER INFORMATION: ORF 23; positive strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (78110)..(76449)
OTHER INFORMATION: ORF 24; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (79864)..(78107)
OTHER INFORMATION: ORF 25; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (81624)..(79861)
OTHER INFORMATION: ORF 26; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (81909)..(81682)
OTHER INFORMATION: ORF 27; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (82346)..(82062)
OTHER INFORMATION: ORF 28; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (82587)..(84446)
OTHER INFORMATION: ORF 29; positive strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (84481)..(85548)
OTHER INFORMATION: ORF 30; positive strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (85556)..(86845)
OTHER INFORMATION: ORF 31; positive strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (87372)..(86803)
OTHER INFORMATION: ORF 32; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (87494)..(88420)
OTHER INFORMATION: ORF 33; positive strandedness; N-terminus only
US-11-205-109-1

Query Match 2.8%; Score 55.2; DB 12; Length 88421;
Best Local Similarity 45.7%; Pred. No. 0.0011;
Matches 275; Conservative 0; Mismatches 318; Indels 9; Gaps 2;
QY 1334 CCAAGCCGCTGGTGTGGCTTCAAGTTCGTGAGCGCTGGTGTCTTCAACAAGCCG 1393
DB 60768 CCGGACGCTCGCCGACCTCGCTCAACTTCAGCTCTCGGTCTTCGAGATCTTCGGC 60827
QY 1394 TCGTGTGGTTCGGCGCGCGCTGCCGGCAAGCAGTACCGCTGTATCAAGCGGACGGCA 1453
DB 60828 CGCTGTGTCCGGCGGCGCAGCATCGAGATCGTCACCGACCTGTCTCGCCCTGGCCGACCCG 60887
QY 1454 TCCCATTCAGAACTACATCGCGCGCACCATCGAGCGGCTGGCGGAGAACTCCACGTGC 1513

```
Db 60888 CCTCCCGCGCTGGAGCGCAGCTGTGTGAGCGGCTGCGTGGCGGTTCTCGCGGCTCC 60947
QY 1514 GCAAGCGAAGTACTTCTTACTACAACTGCTCTACCGGCAAGTTCTCTGCGGACAACTGCC 1573
Db 60948 TCAGCCGGGCGCATCGCGCGCGCACCGCGGAGCGTGTGCTGGCGCGGAGCGCGCTGA 61007
QY 1574 CCACCTTACCTGCGGAGCGGCGCTTCCGCCACCCCTCAAGAGTGGGTGGGACAACTGA 1633
Db 61008 CCGCGGACGTGGTGAACGCCACCGTGCCTGCCCTTCCCGGTGTCCGGGTGGCCAACTCT 61067
QY 1634 CCCTCTTCCACCAATCTTCTTATGAGGAGGTCAAAAGCGCGCACTTACACCAAGGTATTC 1693
Db 61068 ACGGGCGGACCGAGGCGACGCTTACTCGACCGCTTGGCACACCGACCGGAGCGTGACCG 61127
QY 1694 TGATGACCACTGTGAGTGTGCTGTGATGCGCGTGGGCAACG---AGCTGGCGGAGTGCC 1750
Db 61128 GCGGCGCGCGCGCATCGGCGCGCGGTCAACCAACACCCCGCGCTAGTCTCTCGACGACC 61187
QY 1751 TGGCCCAAGCAGGTTGGCGCGGCGGCATCGTCTATCTGGCGCTCCGCTCTCAGCCCGC 1810
Db 61188 GTCTACGCGCGTGGCGCGCGGTGGTGGCGAGCTCTACTGCGCGCGCCAGCTGG 61247
QY 1811 CCTACGCGGAGTGTATCCAGAAGCGGCGCTTCCGAGTGTGCGCTGCATCCGCGCGCACTC 1870
Db 61248 CCGCGGCTATCTGGGCGCGCGCGCTGACCGCGGAGCGCTTCTG-----GSCCTGCC 61301
QY 1871 AGGCTACATGAGCGCGGTCAACATGTACAGCTTCTTACATGGCCCGCGGAGGCG 1930
Db 61302 CGTTGCGCGCGCGGCGGAGCATGTATCCGCAACCGCGCGCGGTCCGGTGAACGCG 61361
QY 1931 CC 1932
Db 61362 AC 61363

RESULT 13
US-11-205-109-1/c
; Sequence 1, Application US/11205109
; Publication No. US20050287641A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Zazopoulos, Emmanuel
; APPLICANT: Staffa, Alfredo
; TITLE OF INVENTION: GENE CLUSTER FOR RAMOPLANIN BIOSYNTHESIS
; FILE REFERENCE: 3002-2US
; CURRENT APPLICATION NUMBER: US/11/205,109
; CURRENT FILING DATE: 2005-08-17
; PRIOR APPLICATION NUMBER: US/09/976,059
; PRIOR FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: US 60/239,924
; PRIOR FILING DATE: 2000-10-13
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 88421
; TYPE: DNA
; ORGANISM: Actinoplanes sp.
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2077)..(3078)
; OTHER INFORMATION: ORF 1; positive strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (3118)..(4032)
; OTHER INFORMATION: ORF 2; positive strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4038)..(5048)
; OTHER INFORMATION: ORF 3; positive strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (6665)..(5814)
; OTHER INFORMATION: ORF 4; negative strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (7703)..(6693)
; OTHER INFORMATION: ORF 5; negative strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (9464)..(8130)
; OTHER INFORMATION: ORF 6; negative strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (9691)..(10761)
; OTHER INFORMATION: ORF 7; positive strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (12751)..(10829)
; OTHER INFORMATION: ORF 8; negative strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (13617)..(12802)
; OTHER INFORMATION: ORF 9; negative strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (15203)..(13614)
; OTHER INFORMATION: ORF 10; negative strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (15591)..(15863)
; OTHER INFORMATION: ORF 11; positive strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (15880)..(19035)
; OTHER INFORMATION: ORF 12; positive strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (19032)..(39713)
; OTHER INFORMATION: ORF 13; positive strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (39713)..(65800)
; OTHER INFORMATION: ORF 14; positive strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (65826)..(66530)
; OTHER INFORMATION: ORF 15; positive strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (66546)..(67370)
; OTHER INFORMATION: ORF 16; positive strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (67384)..(70059)
; OTHER INFORMATION: ORF 17; positive strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (70099)..(70662)
; OTHER INFORMATION: ORF 18; positive strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (70659)..(71906)
; OTHER INFORMATION: ORF 19; positive strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (73439)..(71964)
; OTHER INFORMATION: ORF 20; negative strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (74216)..(73563)
; OTHER INFORMATION: ORF 21; negative strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (75424)..(74213)
; OTHER INFORMATION: ORF 22; negative strandedness
; FEATURE:
```


Db 400 CACGCGCGCATCTCGGGGGGCACATCAACCCGCGTCACTTCGGCCTCTTCCTGGC 459
Qy 1701 CAGGTGGATGCTGATATGCGCGTGGCAAGAGCTGGCGAGTGCCTGSCCAAGCA 1760
Db 460 GCACAAGGTCTCCCTGGTGGCGCGTCTCTATCATCTGTCGCGAGTGCCTCGGCGCCAT 519
Qy 1761 GGTTCGCGCGGGGCATCTGCGCTCGCGCTCTCTAGCCGCGCTTACGCGCA 1820
Db 520 CTGCGGGGTGGCTCTCAAGGGTTCAGAGGGCTCTTTCACAGGTACGGCGGG 579
Qy 1821 GCTGATCAGAAGCGGGCTTCAGCGTGCCTGATCCGCGGCCCACTCAGGGCTACAT 1880
Db 580 CGCAACTGCTCGCTCCGGCTACTCCGCGGCACCGGCTCGGCGCGAGATCATCGG 639
Qy 1881 GGACCGGTCACATGTACAGTCTCTTCTACATGGCCCGCGGAGGGC 1929
Db 640 CACCTTGTGCTGCTTACACCGTCTTCTCCGCCACCGACCCCAAGGCG 688

RESULT 15

US-11-167-048-2

; Sequence 2, Application US/11167048

; Publication No. US20050266487A1

; GENERAL INFORMATION:

; APPLICANT: Batany, Francis

; APPLICANT: Cao, Weiguo

; APPLICANT: Tong, Jie

; TITLE OF INVENTION: HIGH FIDELITY THERMOSTABLE LIGASE AND USES THEREOF

; FILE REFERENCE: 19603/2618

; CURRENT APPLICATION NUMBER: US/11/167,048

; CURRENT FILING DATE: 2005-06-24

; PRIOR APPLICATION NUMBER: 60/106,461

; PRIOR FILING DATE: 1998-10-30

; PRIOR APPLICATION NUMBER: PCT/US99/25437

; PRIOR FILING DATE: 1999-10-29

; PRIOR APPLICATION NUMBER: 09/830,502

; PRIOR FILING DATE: 1999-10-29

; NUMBER OF SEQ ID NOS: 24

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2

; LENGTH: 2025

; TYPE: DNA

; ORGANISM: Thermus sp.

US-11-167-048-2

Query Match 2.8%; Score 54; DB 12; Length 2025;

Best Local Similarity 48.9%; Pred. No. 0.0018;

Matches 175; Conservative 0; Mismatches 180; Indels 3; Gaps 1;

Qy 1468 TACATCGCGCACCATGGAGCGGTGGCGGAGAACTCGCACGTGGCGCAAGCAGAACTAC 1527
Db 907 TACACCGCGCGACCCCGCGCTTCGCCCTCGCCTACAAGTTCCGCGCGGAGGAGAG 966
Qy 1528 TTCTACTACAACTGCCTCACCGCAAGTTCTCGCGGACAACTGCCCCACCTACTCGCG 1587
Db 967 ACCGCGCTCTCTCGTGGCTTCCAGGTGGGGCGGACCGGGCGCATCACCCCGTGGC 1026
Qy 1588 GAGCGGCGCTTGGCCACCTCAAGAGTGGGTGTGACAACTGACCGCTCTCACCAAC 1647
Db 1027 GTTCTGGAGCCGCTCTCATAGAGGCGAGGTGAGCGGGTCACTCCCTCCACACGAG 1086
Qy 1648 TTCTTCATGGAGGAGCTCAAGCGCGCACCTACACCAAGGTATTCTGTATGGACACGTG 1707
Db 1087 AGCTTCATTGAGGAGCTGGAGCTGGCATCGCGACTGGGTG---CTGGTCCACAGGCG 1143
Qy 1708 GACTGGCTGGATATGCCCGTGGCCACAGAGCTGGCGCGAGTGCCTGGCGCAAGAGTTGCG 1767
Db 1144 GCGGGGTGATTCCGAGGTGCTGAGGTCTGAAAGAGCGCGCACCGGGAGGAGAAG 1203
Qy 1768 CCGGGCGGATCGTATCTGGCGCTCGCGCTCCCTCAGCCGCGCCCTAGCGCGAGCTGA 1825
Db 1204 CCATATCTGGCCCGAGAACTGCCCGAGTGGCGGCCACCGCCTCATCAAGAGGGGA 1261

Search completed: March 14, 2006, 01:29:35
Job time : 661 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 14, 2006, 01:18:41 ; Search time 34.5 Seconds
(without alignments)

4959.249 Million cell updates/sec

Title: US-10-620-914-44

Perfect score: 3619

Sequence: 1 aggggtcggctcgtgacgg.....gcccgaagaagacaactaa 1947

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 4886326

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=x1h
-O=/abss/ABSSWEB.spool/US10620914/runat_13032006_101923_5407/app_query.fasta.1
-DB-A Genesecp -QWTF=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0 -LOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abs05h
-USER=US10620914 @CGN 1.1 157 @runat_13032006_101923_5407 -NCPU=6 -ICPU=3
-NO MMAP -NEG SCORES=0 -WAIT -DSRBLCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A Genesecp 21.*

1: Genesecp1980s.*
2: Genesecp1990s.*
3: Genesecp2000s.*
4: Genesecp2001s.*
5: Genesecp2002s.*
6: Genesecp2003as.*
7: Genesecp2003bs.*
8: Genesecp2004s.*
9: Genesecp2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3463	95.7	648	9	ADW72748 Chlamydom
2	891	24.6	752	8	ADP98846 C. albica
3	850.5	23.5	908	9	ADW72753 Neurospor
4	842.5	23.3	845	8	ADR86056 Aspergill
5	462	12.8	19938	6	ABP76680 Streptomy
6	453	12.5	19938	6	ABP76679 Streptomy
7	440.5	12.2	19938	6	ABP76682 Streptomy
8	439	12.1	19938	6	ABP76681 Streptomy
9	437	12.1	19938	6	ABP76679 Streptomy

10	435.5	12.0	19938	6	ABP98398 Streptomy
11	416.5	11.5	19938	6	ABP76678 Streptomy
12	407.5	11.3	19938	6	ABP76682 Streptomy
13	398.5	11.0	19938	6	ABP76678 Streptomy
14	386.5	10.7	1212	2	AAW87503 Human N-m
15	386	10.7	1212	2	AAW87503 Human N-m
16	386	10.7	1212	2	AAW87503 Human N-m
17	386	10.7	1212	2	AAW87503 Human N-m
18	386	10.7	1212	2	AAW87503 Human N-m
19	385	10.6	1061	2	AAW87504 Human N-m
20	384.5	10.6	1081	3	ABP26240 Human N-m
21	384.5	10.6	1081	3	ABP26240 Human N-m
22	384.5	10.6	1081	3	ABP26240 Human N-m
23	384.5	10.6	1081	3	ABP26240 Human N-m
24	381.5	10.5	416	9	ADW72756 Rhodobact
25	379.5	10.5	416	9	ADW72756 Rhodobact
26	379.5	10.5	416	9	ADW72756 Rhodobact
27	378.5	10.5	1615	4	ABP59826 Protein #
28	374.5	10.3	663	7	ABO82097 Pseudomon
29	373	10.3	19938	6	ABP98398 Streptomy
30	372.5	10.3	601	7	ABM86504 Rice abio
31	370	10.2	900	4	ABG03533 Novel hum
32	361.5	10.0	1017	4	ABP59813 TutD prot
33	356	9.8	440	6	ADW48432 Rice prot
34	352.5	9.7	601	7	ABM86504 Rice abio
35	352	9.7	599	4	ABG03530 Novel hum
36	346.5	9.6	1017	4	ABP59813 TutD prot
37	346.5	9.6	1615	4	ABP59826 Protein #
38	333	9.2	1592	4	ABP59827 Protein #
39	332.5	9.2	663	7	ABO82097 Pseudomon
40	332.5	9.2	19938	6	ABP76681 Streptomy
41	332	9.2	536	7	ABO68008 Pseudomon
42	330.5	9.1	1592	4	ABP59827 Protein #
43	329.5	9.1	412	7	ABM87971 Rice abio
44	329.5	9.1	412	7	ABM87926 Rice abio
45	329.5	9.1	1212	2	AAW87503 Human N-m

ALIGNMENTS

RESULT 1

ADW72748

ID ADW72748 standard; protein; 648 AA.

XX

AC ADW72748;

XX

DT 21-APR-2005 (first entry)

XX

DE Chlamydomonas reinhardtii Btal gene, protein.

XX

KW Phospholipid synthesis; transgenic plant; fertilizer; enzyme;

XX

KW betaine lipid.

XX

OS Chlamydomonas reinhardtii.

XX

PN WO2005009115-A2.

XX

PD 03-FEB-2005.

XX

PF 15-JUL-2004; 2004WO-US022789.

XX

PR 16-JUL-2003; 2003US-00620914.

XX

PA (UNMS) UNIV MICHIGAN STATE.

XX

PI Benning C, Riekhof W;

XX

DR WPI; 2005-112975/12.

XX

DR N-PSDB; ADW72746, ADW72747.

XX

PT New composition comprising a purified DNA having an oligonucleotide sequence encoding a protein, useful in producing betaine lipid compounds e.g., Diacylglycerol-O-4'-(N,N,N-trimethyl) homoserine (DGTS).

Db 561 ThrLysValIleLeuMetAspHisValAspTrpLeuAspMetProValAlaAsnGluLeu 580
 QY 1741 GCCGAGTGCCTGCGCAAGCAGGTTGCGCGGGCGGATGCTCATCTGCGCTCC 1800
 Db 581 AlaGluCysLeuAlaLysGlnValAlaProGlyGlyIleValIleTrpArgSerAlaSer 600
 QY 1801 CTCAGCCCGCCCTACCGCAGCTGATCCAGAACGCGGCTTCGACGTGCGCTCATCCGC 1860
 Db 601 LeuSerProProTyrAlaGluLeuIleGlnLysAlaGlyPheAspValArgCysIleArg 620
 QY 1861 CGCGCCACTCAGGCTACATGACCGGCTCAACATGTACAGTCCCTTACATGCGCGCC 1920
 Db 621 ArgAlaThrGlnGlyTyrMetAspArgValAsnMetTyrSerSerPheTyrMetAlaArg 640
 QY 1921 CGAAGCGCGCCCAAGAGGACAAC 1944
 Db 641 ArgLysGlyAlaLysLysAspAsn 648

RESULT 2
 ADP98846
 ID ADP98846 standard; protein; 752 AA.
 AC ADP98846;
 XX 23-SEP-2004 (first entry)
 DT
 DE C. albicans specific gene, orf6.3438, protein sequence.
 XX
 KW Diploid fungal cell; allele; gene disruption cassette;
 KW promoter replacement fragment; antifungal; fungicide; gene therapy;
 KW infection; Candida albicans.
 XX
 OS Candida albicans.
 XX
 PN WO2004056965-A2.
 XX
 PD 08-JUL-2004.
 XX
 PF 19-DEC-2003; 2003WO-US040618.
 XX
 PR 19-DEC-2002; 2002US-0434832P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 PA (ELIT-) ELITRA CANADA LTD.
 XX
 PI Roemer T, Jiang B, Boone C, Bussey H;
 XX
 DR WPI; 2004-500296/47.
 DR N-PSDB; ADP98536.
 XX
 XX Constructing a strain of diploid fungal cells in which both alleles of a
 PT gene are modified comprises modifying the alleles of a gene in the fungal
 PT cells by recombination using a gene disruption cassette and a promoter
 PT replacement fragment.
 XX
 PS Claim 44; SEQ ID NO 7021; 163pp; English.
 XX
 XX The invention relates to a novel method for constructing a strain of
 CC diploid fungal cells in which both alleles of a gene are modified. The
 CC method comprises modifying the alleles of a gene in diploid fungal cells
 CC by recombination using a gene disruption cassette and a promoter
 CC replacement fragment. The invention further comprises: assembling a
 CC collection of diploid fungal cells each of which comprises modified
 CC alleles of a different gene; a strain of diploid fungal cells comprising
 CC modified alleles of a gene, where the first allele of the gene is
 CC inactivated by a gene disruption cassette comprising a nucleotide
 CC sequence encoding an expressible selectable marker; and the expression of
 CC the second allele of the gene is regulated by a heterologous promoter
 CC that is operably linked to the coding region of the second allele of the
 CC gene, and where the gene encodes the polypeptide mentioned above; a
 CC collection of diploid fungal strains comprising the diploid strains cited
 CC above, where substantially all the different genes that encode the above

CC amino acid sequences are modified and are present in different diploid
 CC strains in the collection; a nucleic acid molecule microarray comprising
 CC nucleic acid molecules, where each nucleic acid molecule comprises a
 CC nucleotide sequence that is hybridizable to a target nucleotide sequence
 CC comprising any of the 310 nucleotide sequences listed in the
 CC specification (ADP98516-ADP98825); identifying a gene that is essential
 CC to the survival or growth of a fungus, that contributes to the virulence
 CC and/or pathogenicity of a fungus, or that contributes to the resistance
 CC of a diploid fungus to an antifungal agent; identifying an antifungal
 CC agent that inhibits the growth of a diploid fungus, or a therapeutic
 CC agent for treatment of a mammalian disease; correlating changes in the
 CC levels of proteins or gene transcripts with the inhibition of growth or
 CC proliferation of a diploid fungal cell; a purified or isolated nucleic
 CC acid molecule comprising a nucleotide sequence encoding a gene product
 CC required for proliferation of Candida albicans, where the gene product
 CC consists of any of the above-mentioned amino acid sequences; a vector
 CC comprising a promoter operably linked to the nucleic acid molecule cited
 CC above; a host cell containing the vector; a purified or isolated
 CC polypeptide comprising any of the 61 amino acid sequences given in the
 CC specification (ADP96718-ADP96778); a fusion protein comprising a fragment
 CC of a first polypeptide fused to a second polypeptide, the fragment
 CC consisting of at least 6 consecutive residues of any of ADP98826-ADP99135
 CC ; producing a polypeptide; identifying a compound which modulates the
 CC activity of a gene product encoded by a nucleic acid comprising any of
 CC ADP98516-ADP98825; eliciting an immune response in an animal; a strain of
 CC Candida albicans, where a first allele of a gene comprising any of
 CC ADP98516-ADP98825 is inactive and a second allele of the gene is under
 CC the control of a heterologous promoter; identifying a compound or binding
 CC partner that binds to the polypeptide comprising any of ADP98826-
 CC ADP99135, or its fragment; identifying a compound having the ability to
 CC inhibit growth or proliferation of Candida albicans; inhibiting growth or
 CC proliferation of Candida albicans cells; manufacturing an antimycotic
 CC compound; treating an infection of a subject by Candida albicans;
 CC preventing or containing contamination of an object by Candida albicans,
 CC or for preventing or inhibiting formation on a surface of a biofilm
 CC comprising Candida albicans; a pharmaceutical composition comprising a
 CC therapeutic amount of an agent which reduces the activity or level of a
 CC gene product encoded by a nucleic acid comprising any of ADP98516-
 CC ADP98825 in a pharmaceutical carrier; an antibody preparation which binds
 CC the polypeptide; methods for evaluating a compound against a target gene
 CC product encoded by any of ADP98516-ADP98825; identifying an antimycotic
 CC compound; a computer or a computer readable medium that comprises at
 CC least one of the nucleotide sequences mentioned in the specification or
 CC at least one amino acid sequence selected from ADP98826-ADP99135; a
 CC method assisted by a computer for identifying a putatively essential gene
 CC of a fungus; and a protein array comprising proteins, where at least one
 CC protein comprises an amino acid sequence or a portion of an amino acid
 CC sequence selected from ADP98516-ADP98825. The novel methods and
 CC compositions have fungicide activity. The compositions may be used in
 CC gene therapy. The composition and methods are useful for drug screening
 CC purposes or for diagnosing, preventing or treating infections associated
 CC with Candida albicans. These may also be used for constructing strains
 CC useful for identification and validation of gene products as effective
 CC targets for therapeutic intervention, for identifying and validating gene
 CC products as effective targets for therapeutic intervention, and for
 CC collecting identified essential genes. This sequence represents the
 CC protein of a Candida albicans fungal specific gene of the invention.
 CC NOTE: This sequence was downloaded from an electronic sequence listing
 CC provided on the WIPO website.

XX SQ Sequence 752 AA;

Alignment Scores:

Pred. No.:	1,4e-57	Length:	752
Score:	891.00	Matches:	229
Percent Similarity:	46.6%	Conservative:	98
Best Local Similarity:	32.7%	Mismatches:	232
Query Match:	8	Indels:	142
DB:		Gaps:	21

US-10-620-914-44 (1-1947) x ADP98846 (1-752)

QY 151 CTGGAGAGCTTCTACGGCGCCCGCCGCGCTTCCTT----- 186

Qy	1123	AGCCACAAC	TTCTGGTCCAAGCGCCTCTGGTACTTCCAG-----CACGGCCTG	1170
Db	598	AlapheGln	TyrTrpLeuSerAsnAlaHisIlePheThrAspProAlaGlyArgGlyLeu	617
Qy	1171	TACTACAG	GGGGCGGCATGGCGAAGCTGTGTGGTGTGTGCGTGTGGCCGTGGTGTGCTG	1230
Db	618	TyrAspThr	GlyGlySerArgTyrAlaIleArgPhePheArgTrpIleSerThrLeuPhe	637
Qy	1231	GGACTGG	CAAGACCGTCAAGCGCCTCGCAACGCGCCCAACATGGAGGAGCAGCGCGT	1290
Db	638	PheCysArg	SerAlaValArgLeuLeuSerThrProThrLeuGluGlyGlnArgSer	657
Qy	1291	CTGTGG	GACACCAACATGCTCATCTCGTGAAGAACGCGGCCACGCGCTGGTGTGG	1350
Db	658	IleTyrHis	ThrIlyIle-----ArgProCys-----	666
Qy	1351	CTGTTCGT	CAAGTTCTGTGAGC---CTGGTCTCTTCAACAGGCGGTGTGTGTGGTGGC	1407
Db	667	LeuLeuAsn	ArgPheValAsnGlyLeuValLeuSerSerAspAlaPheLeuTrpSerAla	686
Qy	1408	GGCGGTG	TCGGCGAGCAGTACGGCTCATCAAGGCGGAC-----	1449
Db	687	LeuGlyVal	ProLysAsnGlnValAlaMetIleAlaAspTyrHisArgArgSerIle	706
Qy	1450	-----	-----GGCATCCCATCTGAGAAC	1467
Db	707	SerSerSer	ThrThrProSerSerLysGluLysProSerArgAlaGluAlaIleLeuHis	726
Qy	1468	TACATCG	CGCGCCACCATGAGCGGTGGCGGAGAACTCGCAGTGTGGCAAGCAACTAC	1527
Db	727	TyrThrThr	SerThrLeuAspProValLeuSerThrSerHisLeuAlaSerAspAsnPro	746
Qy	1528	TTCTACT	ACAATGCTCACGGCAAGTCTCTGGCGCAACACTGCCCCACTACCTCGCGC	1587
Db	747	TyrTyrLeu	ValCysValLeuGlyGlnTyrThrArgGlnCysHisProAspTyrLeuSer	766
Qy	1588	GAGCGG	CCCTTGGCCACCCCTCAAGAGT---GGCGTGTGGCAACCTGACCGTCTCCACC	1644
Db	767	ProIleAla	HisSerIleLeuSerAlaProGlyAlaPheAspGlyLeuArgIleHisThr	786
Qy	1645	AACTTCT	TATGGAGAGCTC-----AAAGCGCGCACTACACAAAGGTGATTCTG	1695
Db	787	AspGluIle	GlnGlnValLeuAlaArgPheGlnProGlyThrLeuThrValAlaValVal	806
Qy	1696	ATGACAC	CAGTGGACTGGCTGGATATGCC-----	1725
Db	807	MetAspSer	MetAspTrpPheAspProProSerProGluGluLysGluGlyArgGly	826
Qy	1726	GTGGCCA	ACAGCTGGCGAGTGCCTCGCAAGCAGGTGTGGCGCGGCGGCATCGTCATC	1785
Db	827	LysAlaArg	GluGlnValArgLeuAsnArgAlaLeuLysValGlyGlyLysValLeu	846
Qy	1786	TGGCGCT	CCGCCTCCTCAGCCCGCCCTACGCCGAGCTGTATCCAGAAGCGGGCTTCGAC	1845
Db	847	LeuArgSer	AlaGlyValGluProTyrValArgValPheValGluGluGlyPheGly	866
Qy	1846	GTGCGCT	GTGCATC-----CGCGCGCCCACTAGGCGCTACATGACGCGCGTC	1890
Db	867	AlaArgArg	ValGlyCysArgGluSerGlyArgGlyAspGlnGluCysIleAspArgVal	886
Qy	1891	AACATGT	TACAGCTCTTCTACATGGCCCGCGG	1923
Db	887	AsnMetTyr	AlaSerCysTrpIleLeuGluLys	897
RESULT 4				
ADR86056				
XX	ID	ADR86056 standard; protein; 845 AA.		
XX	AC	ADR86056;		
XX	DT	04-NOV-2004 (first entry)		
XX				

XX Streptomyces viridochromogenes Avi gene cluster polypeptide frame 4.
DE ValValTrpProArgTyrAla***SerAlaAlaAlaSerGlnProArgProProIle 12704
KW Avilamycin; antibacterial; virucide; protozoacide; fungicide; infection;
KW medicine; Staphylococcus aureus; biosynthetic gene cluster; enzyme.
XX
XX Streptomyces viridochromogenes.
XX WO200268436-A1.
XX PN 06-SEP-2002.
XX PD 24-AUG-2001; 2001WO-BP009815.
XX PF 25-FEB-2001; 2001DE-01009166.
XX PR (COMB-) COMBINATURE BIOPHARM AG.
XX PA Weithauer G, Muehlenweg A, Trefzer A, Bechthold A;
XX PI WPI; 2003-018650/01.
XX DR N-PSDB; ABZ37516.
XX DR
XX New avilamycin derivatives, useful for treatment of infections, and
PT nucleic acid encoding avilamycin synthesis enzymes.
XX Example 1; Page 68-301; 319pp; German.
XX
XX The invention relates to avilamycin derivatives (I) with antibacterial,
CC virucide, protozoacide and fungicide activity. (I) are useful for
CC treatment of infections (bacterial, viral, protozoal or fungal), in human
CC or veterinary medicine, particularly where caused by Staphylococcus
CC aureus. (I) are more hydrophilic than known avilamycins. The present
CC sequence is that of an avilamycin synthesis enzyme from the Streptomyces
CC viridochromogenes Avilamycin A biosynthetic gene cluster (ABZ37516)
XX ABZ37516)
XX
XX SQ Sequence 19938 AA;

Alignment Scores:
Pred. No.: 6.77e-25 Length: 19938
Score: 462.00 Matches: 254
Percent Similarity: 32.1% Conservative: 42
Best Local Similarity: 27.6% Mismatches: 257
Query Match: 12.8% Indels: 368
DB: 6 Gaps: 54

US-10-620-914-44 (1-1947) x ABP76680 (1-19938)

QY 1945 AGTTGTCTCTTGGCGCCCTTCC---GGCGGGCATGTAGAGGAGCTGTACATGTTGA 1889
DB 12569 ThrCysAlaGlyArgArgProGlyArgAla***ProAlaArgAlaCys----- 12585

QY 1888 CGCGGTCCATGATACCTCTGAGTGGCGGGCGGATGACGACGACCTCGAAGCCCGCTTCT 1829
DB 12586 ArgGlyGlySerProProArgSerArg---ArgCysArgProArgArgSerCysArgPro 12604

QY 1828 GGA-----TCAGTCGCGGTAGCGGGCTGAGGGAGGCGGAGCGGCAGATGACGA 1778
DB 12605 GlyThr***ArgSerArgArgArgGlySerHisGlyThrArgAlaArgArgArgPro 12624

QY 1777 TGCGCGCGGGCAACTCTGTTGGCCAGGCACTCGGCGAGCTCGTGGCCACGCGGCATAT 1718
DB 12625 AlaArgAlaGlyArgProAlaGlyCysArg***ArgProGlyArgThrProArgArgLeu 12644

QY 1717 CCAGCCAGTCCAGTGGTTCATCAGATCACT----- 1685
DB 12645 ProArgLysProGlySerProProArgSerProArgGlnSerArgProProIlePhe 12664

QY 1685 ----- 1685

DB 12665 ArgProPro***TrpAsnProValHisArgArgGluInPheTyrTrpProProGlyPro 12684

QY 1684 -----TGGTGTAGTGGCGCTTTGAGCT-----CCTCCA--- 1655
DB 12685 ValValTrpProArgTyrAla***SerAlaAlaAlaSerGlnProArgProProIle 12704
QY 1654 -----TGA 1652

DB 12705 ***AlaProSerTyrProThrProCysGlyGlnArgArgGlnileArgSerPheGlnGly 12724
QY 1651 AGAAGTTGGTGGAGACGGTCAGGTTGTCCACCAACGACCTCT----- 1610
DB 12725 AsnSerArgTrpHisArgMetSer---ProProAsnAsnAspArgThrGlyValPhePhe 12743
QY 1609 -----TGAGGTGGCGAAGCGCG----- 1592.
DB 12744 ArgPheSerLeuGlyPheArgAlaGlyTrpArgGlyProPheThrProProSerProAla 12763
QY 1591 -----CCTCGCGCAGGTAGGTGG 1574
DB 12764 AspMetSerTyrGlyValSerAlaPheArgThrHisGlyArgProProSerGlyValPro 12783
QY 1573 GGCAGTTGTGCGCAGCAACTTGCCTGAGGCGAGTTGTAGTAGAGTAGTTCTTCTGCTGC 1514
DB 12784 ***GlyValArgSerGlyAspArgArg-----Cys 12793
QY 1513 GCAGTGGAGTTCTCCGCCACGCCGTCATGGTGGCGG-----CGATGTAGTTCT 1463
DB 12794 Arg***TrpTrpAlaAlaSerArgArgCysTrpProAlaArgProGlyArgGlyAlaSer 12813
QY 1462 CAATGGGATGCGCTCGCTTGTATCAGCGCGTACTGCTTGGCGGCGACCGCGCGCGA 1403
DB 12814 ProAlaGlyGlyGlyProPro-----ProSerProSerArg 12825
QY 1402 ACCACAGCAGCGCTTGTGAAGAGCACCAGGCTCAGAACT-----TGACGAACA 1352
DB 12827 AlaProAlaProPro-----AlaGlyGlyGluArgAlaGlyProTrpGlyArgVal 12843
QY 1351 GCCACACAGCGCTTGGCGCTTCTTACAGAGTGGATGAGCATGCTGCTGCTCCACA 1292
DB 12844 AlaSerProAlaArgSerAlaArgAlaProArgAla-----Pro 12857

QY 1291 GACGGCGCTCTCTCCATTTGTGGCGCGTGGCGAGCGCTTACGCTTGTGCCAGTC 1232
DB 12858 GlyAsnAlaAlaProProAlaThrAspPro-----ArgThrProArg 12872

QY 1231 CCAGCACCGCGCGCAGCACTGCAGCACCAGCAGCTTGGCCATCCCG----- 1181
DB 12873 ProArgProArg-----ArgArgProAlaArgCysProAlaArgProAlaAla 12889

QY 1180 -----CCTGGTGTACAGCGCTGCTGGAAGTACAGAGCGCT----- 1142.
DB 12890 ArgProProGlySerAlaGlyArgAlaGlySerProProGlyValProProArgPro*** 12909
QY 1142 ----- 1142

DB 12910 ProGlyArgAlaArgCysArgArgGlyAspArgAlaArgProArgProSerProPro 12929
QY 1141 -----TGACCAAGAGTGTGGCTGCTGTT--- 1118

DB 12930 ProValAlaCysArgArgArg***SerArgArgArgAlaArgProCysGlyAlaAlaGly 12949
QY 1117 -----GCCACAGGAAGCGCGCGCTTCTTCT---CGTACA 1085
DB 12950 GlyProArgArgArgProAlaArgSerThrSerArgArgProAlaAlaGlyProArgThr 12969

QY 1084 -----GCTCCTCAATGCGGGTGCACGCCCTCCCGCAACAGCTGCC--- 1043
DB 12970 ArgArgAlaArgAlaProGluArgAlaValHisArgAspArgArgThrValAlaArg 12989

QY 1042 -----ACAGTCTCTCAAACTCCAGCTGCTGTAATGCCACT 1007
DB 12990 ArgSerAsnArgAsnArgSerArgValArgProProArgProAlaAlaGlyArgProGly 13009
QY 1006 TCTTCAGCTCCAGGCGCGAGCTGCGCGG---GGTTGCAGTCCACCGACACCTGGC 950

```
Db 13010 ArgGlyArgProValArgAlaGlyAlaArgProGlyProAlaAlaThrAlaProProAsp 13029
Qy 949 CGG-----CCCCCTGCACAGCAGGTTTCAGGGCAT---TGCAGCCCGCTAGTTCAGGG 899
Db 13030 ArgArgGlyProAlaArgProSerGlyProArgHisGlyArgAlaArgArgGlySerGly 13049
Qy 898 ---TGCACACCGTCTCTGGGTTGATCTCCATCACTCCATATCCGGTTCGGGGTCT 842
Db 13050 ArgArgArgProArgProSerGlyArgProAlaAlaProProArgPro---ArgGlyArg 13068
Qy 841 CCCACGACTCGGTACAGGAAGTGGCGCGGACATGGCGCGCGCTCCACCGCTCTCT 782
Db 13069 Pro-----GlySerAlaAlaAspProProSerAlaPro 13079
Qy 781 CGTCAGGCGTGCCCAACGTCGGCAGCGCCCAATCCACAGCTAGTAGGGCGCGCA 722
Db 13080 -----GlyGlySerSerSerCysProGlyGlyArgGly 13090
Qy 721 GCCACGGCAGTAGGAGATCGAACCTGGGTGTCTGCTCCACACGCGCTCCAGCTTCT 662
Db 13091 AlaLeuArgProArgArgProProGlyThr***ProProAlaArgArgProGlyArg 13110
Qy 661 GCTCCAGTAGCGCGCG-----GCTCGGGCGCATCTCA-----TGT 623
Db 13111 SerProArgArg**GlyArgGlyCysProGlyThrGlyGlyCysArgArgProAlaCys 13130
Qy 622 TGT-----CGATGT-----CGAAGATCGATC----- 602
Db 13131 CysAlaThrAlaAlaGlyArgCysProArgArgArgAlaCysProAlaAlaCysGly 13150
Qy 601 GCCAAGAAACGGCGCGCACAG-----GCATCT 572
Db 13151 SerArgArgArgGlyArgAlaArgProLeuAlaProArgArgTrpAsnAlaThrAlaGly 13170
Qy 571 GCGCAGGGCGAGTCTACTTGC----- 548
Db 13171 GlyAlaAlaArgGlyArgAlaAlaGlyLeuArgGlySerArgSerAlaGlyAlaProThr 13190
Qy 547 -----CGCTCAGCTAGAACTCGCAACGCGCACCGCTCTGGCAGCGGTACGAGC 494
Db 13191 SerProArgSerArgArgGlySerGlyArgArgProArgArgArgGlyArgSerGlyPro 13210
Qy 493 AAGCTGTGTGATGAGCTTCT-----GGAAACGGTGAATCATCGTGAGCAGTAGGAGA 440
Db 13211 ArgAlaGly-----CysAlaValAlaSerValArgArgThrArgProProArgArg 13227
Qy 439 AGTGATGAGCTCGCG-----TGCCTCAGGGCGCGCAATTCGCAAGCT 392
Db 13228 ValArgProAlaSerArgGlySerSerAlaArgProArgGlyAlaGlyThrAlaArgArg 13247
Qy 391 CGGCCT-----CCACGACTCGACATCTCTCC----- 365
Db 13248 ThrProGlyGlyArgProArgProProArg**SerArgArgAlaAlaProGlyCys 13267
Qy 364 -----AGCCCTTGGCT-----TCG 350
Db 13268 AlaGlySerProProSerThrArgArgArgSerProTrpProAlaGlyProArgAla 13287
Qy 349 CTTTCT-----TCTTGCCCA-----CTCGCACA 326
Db 13288 ProSerCysArgAlaAlaCysAlaValArgTrpAspSerAlaArgArgProArgAla 13307
Qy 325 GCG-----AGTGCCACAGGTCGACCGTAGTGAAGTCTGAAGT 287
Db 13308 AlaGlyArgAlaGlyGlyProCysAlaGlyCysGlyArgProArg----- 13323
Qy 286 TCGCCAGGTGATGTAATCAGCCATCATATCGCATTTCTCCCGAG-----TGCCAC 236
Db 13324 -----ArgArgProAlaValArgProAlaProAlaProArgSerSerGlyArg 13338
Qy 235 CACCCAGGTCAACCCAGATGAGT-----TCGAGCGCTCG----- 200
```

```
Db 13339 ArgProSerArgProArgArgGlyHisArgArgSerArgTrpThrSerSerArgThrArg 13358
Qy 199 -----CCAGCGGGCAGCAAAAGGCGAGCGGCTGGGGCCCGT 164
Db 13359 TrpProArgProProArgHisCysProArgAlaAlaArgArgProArgArgAlaAsp 13378
Qy 163 AGAAGCTCTCCAGCGCAGCAGCGTATCATCGCCCTTCTTGCTGCCGACACATATGGC 104
Db 13379 ***ProSerProSerArgHisAlaPro---ArgProArgAlaAlaAlaThrTyGly 13397
Qy 103 GCAGAACGGTTCAGGTCTATCTTCATCTGCTGAGCTTGAGCTTCTCCAGGAGAGTTCT 44
Db 13398 -----ArgSer-----AlaGlyProArgSerArg 13406
Qy 43 TCT 41
Db 13407 Ser 13407

RESULT 6
ABP76679
ID ABP76679 standard; protein; 19938 AA.
XX AC ABP76679;
XX DT 26-FEB-2003 (first entry)
XX DE Streptomyces viridochromogenes Avi gene cluster polypeptide frame 3.
XX KW Avilamycin; antibacterial; virucide; protozoicide; fungicide; infection;
XX KW medicine; Staphylococcus aureus; biosynthetic gene cluster; enzyme.
XX OS Streptomyces viridochromogenes.
XX PN WO200268436-A1.
XX PD 06-SEP-2002.
XX PF 24-AUG-2001; 2001WO-EP009815.
XX PR 25-FEB-2001; 2001DE-01009166.
XX PA (COMB-) COMBINATURE BIOPHARM AG.
XX PI Weitnauer G, Muehlenweg A, Trefzer A, Bechthold A;
XX WPI; 2003-018650/01.
XX DR N-PSDB; AB237515.
XX PT New avilamycin derivatives, useful for treatment of infections, and
XX PT nucleic acid encoding avilamycin synthesis enzymes.
XX PS Example 1; Page 68-301; 319pp; German.
XX CC The invention relates to avilamycin derivatives (I) with antibacterial,
XX CC virucide, protozoicide and fungicide activity. (I) are useful for
XX CC treatment of infections (bacterial, viral, protozoal or fungal), in human
XX CC or veterinary medicine, particularly where caused by Staphylococcus
XX CC aureus. (I) are more hydrophilic than known avilamycins. The present
XX CC sequence is that of an avilamycin synthesis enzyme from the Streptomyces
XX CC viridochromogenes Avilamycin A biosynthetic gene cluster (AB237515-
XX CC AB237516)
SQ Sequence 19938 AA;

Alignment Scores:
Pred. No.: 3,19e-24 Length: 19938
Score: 453.00 Matches: 233
Percent Similarity: 30.7% Conservative: 54
Best Local Similarity: 24.9% Mismatches: 289
Query Match: 12.5% Indels: 360
DB: 6 Gaps: 47

US-10-620-914-44 (1-1947) x ABP76679 (1-19938)
```


Db 7215 ProLeuLysArgProAspLeuThrAlaLeuThrAlaArgCysGlyIle***ArgArgSer 7234
 QY 1616 GCGTGGTGGACAACTGACCGTCTCCACCACTTCTTCATGAGAGCTCAAGCGCGCA 1675
 Db 7235 TyrTrpTrp-----AlaGlyLeuAlaArgGlyCysGlyArgLeu 7247
 QY 1676 CCTACACCAAGGTGATCTCATGACACCACTGG-----ACTGGC 1714
 Db 7248 ArgIleProArgProAsnAsnTrpSerArgTrpProIleGluLeuPheSerProMetTyr 7267
 QY 1715 TGATA----- 1720
 Db 7268 TrpIleProSerArgArgProGluTyrArgArgSerGlyLeuSerSerGlyArgProArg 7287
 QY 1721 -----TGCCCTGGCCCAACGAGCTGGCCGAGTGCCTGGCCCAACGAGTTCGCG 1768
 Db 7288 ArgArgSerArgLeuProTrpGluThrAlaArgSerSerAlaTrpThrSerProThrSer 7307
 QY 1769 CGGGCGGCATGTCATCTGGGGTCCGCTCCCTCAGCCCGCCCTAGCCGAGCTGATCC 1828
 Db 7308 SerArgThrProArgAlaGlyTyrProThrProCysSerCysProValThrSer---Ser 7326
 QY 1829 AGAAGCGCGCTTCGAGTCC----- 1849
 Db 7327 AlaThrThrThrSerSerCysThrTrpAlaThrThrThrSerTrpThrAlaSerSerThr 7346
 QY 1850 -----GTCATCGCCGCGCCACTCAG----- 1873
 Db 7347 SerTrpArgThrSerAlaAlaThrSerProProArgArgSerCysSerProGlySerPro 7366
 QY 1873 ----- 1873
 Db 7367 ThrArgAlaGlySerAlaTrpProAsnLeuThrArgThrAlaGly***ProArgSerTrp 7386
 QY 1874 -----GTCATCGCCGCGCTCAACATGTACA-----GCT 1903
 Db 7387 ArgSerProArgThrProGlyAlaThrTrpArgSerSerAlaCysThrProSerAlaArg 7406
 QY 1904 CTTCTTACA-----TGCCCGCGCGAAGGCGCCCAAGAGGCAACT 1945
 Db 7407 ProCysThrArgProTrpProAlaSerGlyArgProGlyGlyThrSer 7422
 RESULT 7
 ID ABP76682
 DE ABP76682 standard; protein; 19938 AA.
 AC ABP76682;
 XX
 XX 26-FEB-2003 (first entry)
 DT
 XX Streptomyces viridochromogenes Avi gene cluster polypeptide frame 6.
 DE
 DE Streptomyces viridochromogenes
 KW Avilamycin; antibacterial; virucide; protozoicide; fungicide; infection;
 KW medicine; Staphylococcus aureus; biosynthetic gene cluster; enzyme.
 XX
 XX Streptomyces viridochromogenes.
 XX
 XX WO200268436-A1.
 XX
 XX 06-SEP-2002.
 PD
 XX 24-AUG-2001; 2001WO-EP009815.
 PF
 XX 25-FEB-2001; 2001DE-01009166.
 PR
 XX (COMB-) COMBINATURE BIOPHARM AG.
 XX
 XX Weithauer G, Muehlenweg A, Trefzer A, Bechthold A;
 XX
 XX WPI; 2003-018650/01.
 DR N-PSDB; ABZ37516.
 DR
 XX

PT New avilamycin derivatives, useful for treatment of infections, and
 XX nucleic acid encoding avilamycin synthesis enzymes.
 PS Example 1; Page 68-301; 319pp; German.
 XX
 CC The invention relates to avilamycin derivatives (I) with antibacterial,
 CC virucide, protozoicide and fungicide activity. (I) are useful for
 CC treatment of infections (bacterial, viral, protozoal or fungal), in human
 CC or veterinary medicine, particularly where caused by Staphylococcus
 CC aureus. (I) are more hydrophilic than known avilamycins. The present
 CC sequence is that of an avilamycin synthesis enzyme from the Streptomyces
 CC viridochromogenes Avilamycin A biosynthetic gene cluster (ABZ37515-
 CC ABZ37516)
 XX
 SQ Sequence 19938 AA;
 Alignment Scores:
 Pred. No.: 2,75e-23 Length: 19938
 Score: 440.50 Matches: 182
 Percent Similarity: 38.9% Conservative: 60
 Best Local Similarity: 29.3% Mismatches: 235
 Query Match: 12.2% Indels: 145
 DB: 6 Gaps: 29
 US-10-620-914-44 (1-1947) x ABP76682 (1-19938)
 QY 155 AGAGCTTCTACGGGCCCCAGCGCGCTTCTGCTGCGCGCTGGCGGAGCGCTCGAAC 214
 Db 5169 ArgArgSerIleSerGlnLeuProValProAspValHisPheTyrProGlnCysAlaAla 5188
 QY 215 TCATCTGGGTGACCTGGGTGGTGGCACCTGGGGAGAATGCGATATGATGGCTGATTACA 274
 Db 5189 Thr***AsnLeuAsnTyr----- 5194
 QY 275 TCGACCTCGAGGAGTCAAGTCCATCTACGTGCTGCACCTGTGCCACT-----CGCTGT 328
 Db 5195 ---GlnTrpGlyAlaProCysProLysLysCysSer***SerAlaValGlyArgArgAsp 5213
 QY 329 GCGAGTGGCCCAAGAGAAGAGGCGCAAGGCTGGAGAATGTCCAGGTGCTGGAGG 388
 Db 5214 ProArgArgProCysSerAlaAlaArgProAlaCysPro***SerTyrTrpAsnGlyArg 5233
 QY 389 CGACCTGTCGCAATTTGGCCCTTGAGGGGACCGGAGGCTCATCACCCTTCTCTACT 448
 Db 5234 ProSerArgAlaThrThrSerAlaSerArgSerProProArgAlaAlaProSerSerThr 5253
 QY 449 CGCTCAGCATGATCCACCGTCCACACGTATCGACCGAGGCTGTCTGCTACTGTGCC 508
 Db 5254 SerSerAlaArgTrpArgSerThrProAlaValThrArgSerArgThrAlaCys--- 5272
 QY 509 AAGACGCGCTGGTGGCGGTTGCGGACTTCTACGTGAGCGCAAGTACGACCTGCCCTGC 568
 Db 5273 -----CysCysAlaGly 5276
 QY 569 GCGAGATGCCGTGTCGCGCGCTTCTTCTGGGATCGATCTTCGACATCGACACATTG 628
 Db 5277 AlaArgArgThrGlyArg-----SerThrGlyArgArgSerSerValPro----- 5291
 QY 629 ACATCGGCGCGCGCGCGCTACCTCGAGCAGAGAGTGGAGCGGTGTGGAGCAGA 688
 Db 5292 ThrSerAlaProGlyArgSerThr-----AlaThrThrSerThrCysSer***Arg 5309
 QY 689 ACACCCAGGGTTTCGATCCCTACGTGCGCGCTGCGCGCTTCTACTACTAGTGTGGATTG 748
 Db 5310 ThrArgSerSerArgAlaProThr-----SerProArgAlaSerArgSerSer 5325
 QY 749 GCGCGCTGCCAGCGGTGGCCAGCCCTGCACGAGGAGCGGTGGAGCGGCCCGCCATGT 808
 Db 5326 AlaCysCysSerThrAlaThrAlaArgLeuGlyArgSerGlyTrpThrProThrProVal 5345
 QY 809 TCCCGCCACCTTCTCTGTACACGCGACTGCTGGGAGGAGCCCGGCGGATATGGAGTGA 868
 Db 5346 Ser-----TrpSerAlaArgSerSerThrThrSer----- 5355

```
QY 869 TGGAGATCAACCCCAAGGACACGCTGCTGACCTGACTAGCGGGCTGCAATGCCCTGA 928
Db |||||SerThrProAla-----5366
QY 929 ACCTGCTGTGCAGGGGGCGCGCAGGTGGTGTGCTGGACTGCAACCCCGCAGTCGG 988
Db |||||-----GlyProAla-----5369
QY 989 CGCTTCTGGAGCTGAAGAGTGGCCATTGACGAGCTGGAGTTGAGGAGCTGGCAGC 1048
Db |||||-----ThrAlaAlaArgThrArgSerSer 5378
QY 1049 TGTTCGCGGAGG-----GGTGCACCGCGCATTTGAGGAGCTGACGAGAAGCTGG 1102
Db |||||-----GlyAlaThrGlyArgAlaAlaArgCysCysProValArgProGln 5398
QY 1103 CGCCCTTCTGTGCGCAACAGCCCAACTTCTGGTCCAAAGCGCCTCT-----1150
Db |||||-----GlyAlaSerThrSerAlaProArgThrAlaGlyThrGly***SerArgCysAlaGly 5418
QY 1151 ---GGTACTTCCAGCAGCGCTGTACTACGAGCGCGCATGGGCAAGCTGTGCTGGTGC 1207
Db |||||-----ThrGlyThrAlaSerValSerSerAlaThrArgThrGlySerThrAlaAlaGlyAsn 5438
QY 1208 TGCAGTCCCTGGCGTGTCTGGAC---TGGGCAAGCCGTCGAGCGCTGCCCAACG 1264
Db |||||-----ThrProThrTrpArgArgCysTrpProArgTrpSerArg-----AsnProArgProCys 5456
QY 1265 CGCCCAATCGGAGGAGCAGCGCTGTGTGGGACGACCAATGCTCATCCTCATCTCGTGA 1324
Db |||||-----AlaThrSerTrpArgThrAlaSerThrSerArgAlaTrpGlyTrpSerArgThrSer---5475
QY 1325 AGAAGCGGCCAGCCCTGGTGTGGCTGTTCGTCAGTTGCTGAGCTCG-----TGC 1378
Db |||||-----ProThrSerProThrAlaSerThrAlaProAlaThrSerProAlaThrProProAla 5495
QY 1379 TCTTCACAGGCGCTGTGGTTCGGCGCGCGTGGCGGCAAGCAGTACGCGTGA 1438
Db |||||-----ArgArgArgProGlyArgSerThrSerArgSerThrAlaThrProThrSer 5529
QY 1556 TCCTGCGGACAACTGCCCACTACCTGCGGAGCGGCTTCGCCACCTCAAGAGTG 1615
Db |||||-----GlyCysSerProThrCysProAlaCysThrSerAsnArgProAlaArgProProThrSer 5565
QY 1616 GCGTGTGGACAACTGACCG-----TCTCCACCAACTTCTTCATGAGGAGCTCAAG 1669
Db |||||-----SerTrpProThrArg***ProArgSerAlaProArgProSerThrSerArgSerThrAla 5585
QY 1670 CGCGCACTACCAAGGTGATTCGTATGACACAGTGGACTGCTGATATGCCCTGG 1729
Db |||||-----ArgTrpProSerPro-----AsnTrpSerArgAspTrp 5596
QY 1730 CCAACGAGC---TGGCGAGTGCCTGCGCAAGCAGTTGCGCGGCGCATCG---TCA 1783
Db |||||-----ProThrSerThrProSer-----ProAlaArgThrArgThrAlaProSerArgArg 5614
QY 1784 TCTGCGCTCGCCCTCCCTCAGCCCGGCTCAGCGGAGCTGATCC-----1828
Db |||||-----ThrProArgProThrProCysArgArgThrThrProSerAlaSerCysSerArgArgPro 5634
QY 1829 -----AGAGCGGGCTTCGACGTGGCTGCTGCTGCTGCGCGGCCACTCAGG 1873
Db |||||-----ArgThrProVal***TrpArgArgProGlyCysProProHisArgSerSerAlaArgLeuPro 5654
```

```
QY 1874 GCTACATGACCGCGTCAACATGTACAGCTCCT-----TCTACATGG-----CCGCGC 1921
Db |||||-----AlaArgTrpThrProThrThrSerThrThrProProArgValSerThrTrp***ProAla 5674
QY 1922 GGAAGG 1927
Db |||||-----ArgArg 5676
RESULT 8
ABP76681
ID ABP76681 standard; protein; 19938 AA.
XX AC ABP76681;
XX XX 26-FEB-2003 (first entry)
XX DE Streptomyces viridochromogenes Avi gene cluster polypeptide frame 5.
XX KW Avilamycin; antibacterial; virucide; protozoacide; fungicide; infection;
XX KM medicine; Staphylococcus aureus; biosynthetic gene cluster; enzyme.
XX OS Streptomyces viridochromogenes.
XX XX WO200268436-A1.
XX PN 06-SEP-2002.
XX PD 24-AUG-2001; 2001WO-EP009815.
XX PF 25-FEB-2001; 2001DE-01009166.
XX PR (COMB-) COMBINATURE BIOPHARM AG.
XX PA Weitnauer G, Muehlenweg A, Trefzer A, Bechthold A;
XX XX WPI; 2003-018650/01.
XX DR N-PSDB; ABZ37516.
XX XX New avilamycin derivatives, useful for treatment of infections, and
XX PT nucleic acid encoding avilamycin synthesis enzymes.
XX PS Example 1; Page 68-301; 319pp; German.
XX CC The invention relates to avilamycin derivatives (I) with antibacterial,
XX CC virucide, protozoacide and fungicide activity. (I) are useful for
XX CC treatment of infections (bacterial, viral, protozoal or fungal), in human
XX CC or veterinary medicine, particularly where caused by Staphylococcus
XX CC aureus. (I) are more hydrophilic than known avilamycins. The present
XX CC sequence is that of an avilamycin synthesis enzyme from the Streptomyces
XX CC viridochromogenes Avilamycin A biosynthetic gene cluster (ABZ37516)
XX CC ABZ37516)
SQ Sequence 19938 AA;
Alignment Scores:
Pred. No.: 3,57e-23 Length: 19938
Score: 439,00 Matches: 258
Percent Similarity: 30,6% Conservative: 39
Best Local Similarity: 26,6% Mismatches: 257
Query Match: 12,1% Indels: 416
DB: 6 Gaps: 57
US-10-620-914-44 (1-1947) x ABP76681 (1-19938)
QY 1945 AGTTGCTCTCTGCGCCCTTCGCGGCGCATGTAGAGGAGCTGTACATGTTGACGC 1886
Db |||||-----AlaCysProAlaTrpLeuProPheProPro-----1949
QY 1885 GGTCCATGTAGCCTGAGTGG-----1865
Db |||||-----GlyProAlaHisProLeuTrpLeuAlaProLeuArgThrProAlaArgProAlaGlyLeu 1969
QY 1864 -----CGCGGC 1859
```

1970 LeuProProLeuAlaArgProArgPro***AlaProArgValProLeuProValArgPro 1989
1858 GGATGC-----AGCGCAGT---CGAAGCCGCT-----1832
1990 GlyThrArgThrProArgProSerAlaArgSerArgProProGlnArgSerArg 2009
1831 ---TCTGGATCAGCTCGCGTAGGCGGCTGAGGAGG-----1796
2010 AlaAlaGlyAlaProSerArgThrAlaProAspAlaArgThrProProCysAlaSerPro 2029
1795 -----CGAGCGCCAGATGATGCGCGCGCGCAA---CCTGCTGG---CCA 1751
2030 ***SerCysArgAlaArgProArgGlyArgProProArgProProCysThrAlaPro 2049
1750 GGCACCTCGGCA---GCTCGTTGGCCACGCGGCATATCCAGCCAGTCCAGCTGGT-----1700
2050 ProThrArgProCysAlaGlyAlaProGlyValArg**GlyArgProAlaGlyCysGly 2069
1699 -----CCATCAGATCACTTGTGTAGTGGCGCTTTGAGCT 1661
2070 AlaAspArgProValArgProProCysGlyThrProTrp-----Ser 2083
1660 CCTCCATGAAGATTGGTGGAGACGCTCAGTTGTCCACCGCCACTCTTGAGGGTGG 1601
2084 ProProArgArgAlaArg-g-----Trp 2092
1600 CGAAGCGCGCTCGCGCAGCTAGTGGCGCAGTTGT---CGCGCAGGAATTCGCGGTGA 1544
2093 ProProProAlaArgAlaGly-----GlySerAlaProGlyCysGlyThrProArgArg 2110
1543 GGCAGT-----TGATGAGAGTGTCTGTGGCCAGCTGCGAGTCTCCGCCA---1493
2111 GlySerProTrpAlaTrpArgSerGlyThrAlaArgGlyArg---SerAlaProProThr 2129
1492 ---CGCGTCCATGGTGGCGCGCATGT-----AGTTCTCAATGGGATCGCTCCGCT 1442
2130 ArgArgArgProTrpSerGlyGlySerGlyAspSerSerArgSerGlyProArgPro---2148
1441 TGATCAGCGGTACTGTTCCCGGCACGCGCGCGCCAGCCAGCAGCGCTGTGTA 1382
2149 -----ArgAlaGlyCysProProArgArgProArgAlaProGlySerProAlaAla 2165
1381 AGA-----GCACAGGCTCAGAACTTGACGACAGCC 1349
2166 ArgArgArgTrp***ProArgArgArgProAlaProSerSerArgThrArg-----2183
1348 ACACGAGCGCTTGGGCGCTTCTTACGAGTGGATGACATGTGCTGT-----1298
2184 ---ProAlaGlyArgGlyArgAlaAlaValSerArgAlaAlaHisCysAlaSerArgSer 2202
1297 -----CCCACAGACGCGCTGCTCCT 1277
2203 ProGlySerSer***SerAlaArgTrpProAspArgProPro***SerGlySerAlaPro 2222
1276 ---CCATTGGGCG-----CGTTGGCAGGCGCTTGACGCTTGGCCAGTCCCA 1229
2223 ArgArgLeuGlyAlaArgValArgTrpProGlySer-----SerCysArgAlaPro 2240
1228 GCA-----CCACGCGCAGGC 1214
2241 SerArgArgArgGlySerSerGlyAlaGlyArgGlyArgSerGlyCysArgArgProGly 2260
1213 ACTGCAGCA-----CCACGACAGCTTGC-----1190
2261 SerArgAlaValArgProGlyGlyAlaCysArgArgAlaProGlySer***ProAlaArg 2280
1189 CCATGCCGCGCTGTAGTACAGCCGT-----GCTGGAGTACAGAGCGCTTGGACC 1136
2281 ProProArgProAlaGlyThrProArgProProArgGlyArgGlyProArgAla 2300
1135 AGAAGTTGGCTGGTTGGCAGAGAGGCGCCAGCTTCTCTGT-----1088

2301 ArgArgAlaGly---ArgSerThrArgSerAlaAlaArgSerArgProGlyArgPro 2319
1087 ---ACAGCTCCTCAATGCGCGGTGCACGCT-----CGCCGACAGCTGCCACA 1040
2320 ProArgSerProGlyGlyAlaAlaGlyArgProArgProGlyArgAlaAlaValAlaHis 2339
1039 CGTCTCAAACTCCAGTCTGTGAA-----TGGCCACCTTCTTTCAGCT 998
2340 ArgArgGlyArgProAlaArgSerValArgGlySerProProArgProProAlaProVal 2359
997 CCAGAAGCGCGACTCGCG---GTTGTCAGTCCACCGACACACCTGCGCGCCCT 941
2360 ProValArgProAlaProArgThrAlaCysAlaGlyLeuProAla---ProPro 2378
940 GCACGACGAGTTTCAGGCGCATTCGACGCGCTAGTCAAGGTCAGCAGCAGCTGTCTTGG 881
2379 ArgProAlaAlaGly-----ArgAlaSerSerAlaProArgPro---2392
880 GGTGATCTCCATCACCCTCCATATCCGCTCGGGTCTCCACGACTGGCTGTACAGGA 821
2393 ---ProArgArgProGlyValProArgGlyProProProProAlaAlaAla 2409
820 ---AGTGGCGCGCAACATGGCG-----GCC 797
2410 ArgArgProValArgTrpSerGlyAlaAlaCysProArgArgCysValaArgCys 2429
796 GCTCCACGCGCTCCTCGTGCAGGCGTGCACACGCTGGCGCGCCCAATCCACAGT 737
2430 CysProValGlnProArgArgSerAlaGlyArgArgProProSerGlyArgArgSerAla 2449
736 AGTAGG-----GGCGCGCAGCCAGCGCAGTAGGAGGATCG-----701
2450 AlaArgProArgGlyAlaGlyAlaAlaGlyThrSerArgArgAlaProProGlyArgPro 2469
700 ---AACCTGGGTCTCTCTCC-----680
2470 SerGlyThrArgProSerProProProGlyAlaAlaCysProArg***GlyProThr 2489
680 -----680
2490 AlaGlyProProGly***ProProAlaArgAlaGlyProValAlaHisGlySerValPro 2509
679 ---ACACGCGCTCCAGTCTCTGCTCCAGGT 653
2510 GlyProProValArgProArgArgValArgAspArgArgProProAlaProPro 2529
652 AGCGCGCGCTCGG-----638
2530 ArgGluAlaAlaArgProGlyProProProProAlaHisSerAlaProSerArgProArg 2549
637 ---GGCGGATGT 629
2550 ProSerHisArgValProAlaSerCysHisProGlyGlyArgGlyLeuAspGlyTrpCys 2569
628 CAATGTTGT-----620
2570 AlaGlyCysProAlaThrLeuThrIleGlyArgArgAlaHisThrCysProLeuTrpVal 2589
619 ---CGATGTCGAGATCGATCGCCAGAGAAC 590
2590 GlySerArgGlyAlaLeuGlyGlnAlaAspArgLeuLysAlaGlyIleGlyArgLeuAsn 2609
589 GGC-----587
2610 GlyProValArgSerTrpProAlaProAlaValProAspThrArgArgSerArgAlaAla 2629
586 ---GCACCGCGCATCTGCGCGCAGCGGCGAGT---557
2630 Ser***ProThrThrGlyProGlyArgAlaGlyArgAspArgGlyProGlyArgGlyVala 2649
556 ---CGTACTTGGCGCTCA---542
2650 ***ArgSerValPro***SerArgProProValProAlaTrpArgArgCysArgAlaLeu 2669

QY 541 -----CGTAGACTCGGCAACCCACACGCGCGCTCTTGGGACAGGTAGC-----ACCAAG 491
 Db 2670 LeuProArgArgSer---AlaLeuProProTrrLysArgGly**GlyAlaAspSerAla 2688
 QY 490 CTTGGTGCATGACGTTGTGAACCGGTGAATCATCGTGACGAGT-----446
 Db 2689 ProProAsnLeuPheCysAlaAlaGlyGlyLeuSerArgSerProSerProTrrPrrSer 2708
 QY 445 -----AGGAGAAGGTGATGA 431
 Db 2709 AlaSerProThrSer**CysArgCysAlaProAspArgGlyAspArgAlaArgGln** 2728
 QY 430 GCCTCGCGTGCCTCAGGGGGCGCAATTTGGACGGTCCGCTCCA---CGACTTGA 374
 Db 2729 ProArgArgArgProThrAlaAlaAlaAlaGlySerArgProProAlaArgProGly 2748
 QY 373 CATTCTTCAGCCCTTGGCCTTCGCTTCT-----344
 Db 2749 -----SerAlaProProProSerProGlyArgArgAlaGlySerArgArgArgArg 2766
 QY 343 ---TCTTGGCCACTCGCACGAGGTGGCACAGGTGCACCGTACAGTACTGAACT 287
 Db 2767 ArgSerArgGlySerGlySerProSerGlyThrGlyArgCysProCysTrrProArgPro 2786
 QY 286 TCGCCA-----GGTCATGTATCAG---CCATCATATCGACATTTCTCCCGATGC 239
 Db 2787 ValProSerArgThrAlaArgCysGlyArgAsnProAlaArgSerAlaPro**Cys 2806
 QY 238 CAC-----CACCCAGGTCAACCCAGATGAGTTTCGAGCGCTCGG-----200
 Db 2807 ArgSerArgArgProArgProArgAlaProArgThrGlyProProGlyArgThrSerPhe 2826
 QY 199 -----CCAGCGGCGCAAAAGGCGAGCGCTCGGCGCCGTAGAACG 158
 Db 2827 ArgCysThrProArgSerProTrrSerGlnGlySerAlaSerAlaSerValArgGlu 2846
 QY 157 TCTCCAGCGCAGACGCTGATCGCCCT 128
 Db 2847 AlaProGlyArgArgAlaArgHisLeuPro 2856

RESULT 9
 ABP76679
 ID ABP76679 standard; protein; 19938 AA.
 XX AC ABP76679;
 XX DT 26-FEB-2003 (first entry)
 XX DE Streptomyces viridochromogenes Avi gene cluster polypeptide frame 3.
 XX KW Avilamycin; antibacterial; virucide; protozoacide; fungicide; infection;
 XX KW medicine; Staphylococcus aureus; biosynthetic gene cluster; enzyme.
 XX OS Streptomyces viridochromogenes.
 XX PN WO200268436-A1.
 XX PD 06-SEP-2002.
 XX PF 24-AUG-2001; 2001WO-EP009815.
 XX PR 25-FEB-2001; 2001DE-01009166.
 XX PA (COMB-) COMBINATURE BIOPHARM AG.
 XX PI Weitnauer G, Muehlenweg A, Trefzer A, Bechtold A;
 XX DR WPT; 2003-018650/01.
 XX DR N-PSDB; ABZ37515.
 XX PT New avilamycin derivatives, useful for treatment of infections, and
 nucleic acid encoding avilamycin synthesis enzymes.

XX Example 1; Page 68-301; 319pp; German.
 XX The invention relates to avilamycin derivatives (I) with antibacterial,
 CC virucide, protozoacide and fungicide activity. (I) are useful for
 CC treatment of infections (bacterial, viral, protozoal or fungal), in human
 CC or veterinary medicine, particularly where caused by Staphylococcus
 CC aureus. (I) are more hydrophilic than known avilamycins. The present
 CC sequence is that of an avilamycin synthesis enzyme from the Streptomyces
 CC viridochromogenes Avilamycin A biosynthetic gene cluster (ABZ37515-
 CC ABZ37516)
 XX SQ Sequence 19938 AA;
 Alignment Scores: 5.03e-23 Length: 19938
 Pred. No.: 437.00 Matches: 237
 Score: 32.0% Conservative: 43
 Percent Similarity: 27.1% Mismatches: 271
 Best Local Similarity: 12.1% Indels: 324
 Query Match: 6 Gaps: 51
 DB:
 US-10-620-914-44 (1-1947) x ABP76679 (1-19938)
 QY 1945 AGTTGTCTTCTTGGCGCCTTCGCGGCGCATGTAGA-----1907
 Db 6293 SerThrSerGlyTrrArgPro-----GlyProCysArgGlyProSerValArgThrPro 6310
 QY 1906 -----AGGAGCTGT-----1898
 Db 6311 AlaGlyArgLeuCysProValSerAlaThrAlaAlaProTrrProMetTrrSerSerAsn 6330
 QY 1897 -----ACATCTTTCGCGCGT 1893
 Db 6331 ArgArgProProArgProProAlaProArgArgSerArgProGluProCysSerArgCys 6350
 QY 1882 CCATGTAGCCCTCAGTGGCGCGCGCATGACGAGCGCGCATCGAAGCCGCTTCTGGATCA 1823
 Db 6351 ProArgProProArg---ArgArgSerValThrAlaArgProArgSerProSerGlySer 6369
 QY 1822 GCTCGCGGT-----AGGCG 1808
 Db 6370 ArgArgAlaLeuThrTrrProArgSerGlyThrProTrrProTrrProTrrProTrrPro 6389
 QY 1807 GGCTGAGGAGCGGCGCGCGCATGACGATGCGCGCGCAACCTGCTGCGCCAGGC 1748
 Db 6390 SerThrGlyArgProSer-----TrrArgProAlaGlyThrAsnTrrPro---6404
 QY 1747 ACTCGGCA-----GCTCGTGGCCACGCGGCATATCCAGCCAGTCCACGTGGTCCATCA 1694
 Db 6405 -----ProGlyCysAlaArgTrrProLeuArgAsnProArg-----ProVal 6418
 QY 1693 GAATCACTTGGTGTAGTGGCGCTTTGAGCT-----1661
 Db 6419 SerSerProGlyProArgCysArgMetProAlaValProSerGlyCysSerArgValThr 6438
 QY 1660 ---CCTCCATGAAGATTGG-----TGGAGCGGTGAGTGTCCACACGCG 1616
 Db 6439 GlyProSerGlyProGlyTrrAlaAlaAsnCysTrrLysProSerArgSerProArg 6458
 QY 1615 CACTCTTGGGTGGCGAAGCGCGCTCGCGCAGTAGTGGGCGCAGTTGTGCGCGCAGGA 1556
 Db 6459 ***SerThrAsnTrrSerArgCysSerArg-----ArgArgSerGlySerArgProGly 6476
 QY 1555 ACTTGGCGGTGAGGCGAGTTGTAGTAGAAGTAGTTCTCTCTCGCGCAGTGGGAGTTCTCCG 1496
 Db 6477 ArgCys-----CystTrrArgAlaThrThrPro 6485
 QY 1495 -----CCACGCGGTCCATGTTGGCGCGCATGTAGTTCT-----1463
 Db 6486 ArgSerThrAlaProArgPro***SerSerArgCysSerSerGlySerProArgCysGly 6505
 QY 1462 ---CAATGGGATGCCGCTCGCTTGTATCAGCGGTACTGTTCCCGGCGCACGCGCGCGC 1406

6506 GlyHisGlyAlaSerSerProPro- ||||| ProSerSerAlaThr 6518
1405 CGAACACACACACGGCTTGTGAAGAGACACAGGCTTCAGAACTTGACCAACAGGCACA 1346
6519 ArgSerValArgSerProProSerProProGly- ||||| ArgProProPro 6534
1345 CCAGCGCTTGGCCCGCTTCTTCCAGAAAGTGATGACATGTTGCTGTCTCCACAGACGGC 1286
6535 ProThrGlyProGlySerAlaValGly- ArgCysCysCys 6548
1285 GCTGCTCTCATTTGGCGCGCTTGGCAGGCGCTTGACGGCTTGCCTCCAGTCCACGCA 1226
6549 AlaArgProArgGlyValAlaArgTrpArgTrpSer- ValCysProSerThrArg 6566
1225 CCAGCGCCAGCCTCCAGCA- SerAlaValGly- ArgCysCysCys 1205
6567 ProProSerGlySerArgAlaMetThrArgTrpSerArgProSerArgProArgArgArg 6586
1204 -----CCAGCAGACCTGCCCTCCCGCTGTAGTACAGCGCGTGTGGAAGTACC 1151
6587 ProProSerProValThrProAlaArgSerArgArgSerAlaAlaGlyProThr 6606
1150 AGAGCC-----GCTTGACC 1136
6607 ArgGlyTrpSerTyrglyGlySerProProThrTrpProSerThrAlaArgTrpThr 6626
1135 -----AGAAAGTTGGCTGTTTGCAGAGGAAGGCGCCAGCT- 1097
6627 ProCysSerThrGlyCysAlaArgProProThrSer***ValProAlaHisArgThrArg 6646
1096 ---TCTTCTCGTACAGCTCCTCAATCGCGGTCACGC- 1061
6647 ArgSerThrArgArgSerArgThrArgGlyProArg***ProProThrAlaArgThr 6666
1060 CTTGCGGGAACAGCTGCCACACT- CTTCAAACTCCA 1025
6667 GlyArgArgThrCysAlaThrArgCysGlySerProProArgSerProArgProArg 6686
1024 GCTGCTGAATGGCCACTTCTCAGCTCCAGAGCGCGACTCG- 980
6687 ThrAlaThrGlyArgSerSerCysProArgThrProTrpSerArgThrArgSerThr 6706
979 -----CGGGTTGCAGT- 968
6707 ArgArgTrpProAlaAlaAlaTrpArgThrCysSerSerAspArgArgTyrglyAlaThr 6726
967 -----CCACGACACACCTGGCGGCGCCCT- GCACGACGA 932
6727 SerProArgProGlyProSerAlaProProTrpAlaProProThrAlaThrAla***Ala 6746
931 GGTTCAGGCGATTGACGCGCGCTAGTCAGGCTCAGCAGCGCTGCTGGGTGATCT 872
6747 TrpThrGlyArgSerSerArgThrGlyThrSerLysSerCys- Arg 6762
871 CCATCACCT-----CCATATCCGCTCGGGTCTCTCCACAG 836
6763 ProThrProGlySerThrAlaArgCysGlyValProSerProGlyArgGluArgProSer 6782
835 ACTGCGGTACAGAGGTGGCGGGAACATGGCG- CCGCTCCACGCGCTCCT 782
6783 AlaAlaThrThrSerThrProThrArgCysTrpAlaArgProAlaSerArgAlaAla 6802
781 CGTGCAGGCGGTGC- CAACCTCGGCGGCGCAATCCACAGTAGTAGGCGCGC 725
6803 ThrCysGlyCysGlyThrAlaArgTrpThr- ThrThrAlaAlaArgThrArg 6819
724 GCAGCC-----ACGCGACCTAGGGATCGAACCCTGGGTG- 689
6820 AlaAlaThrProSerThrAlaTrpArgSerArgProProCysTrpArgSerArgSer 6839
688 -----TCTGCTCCACAGCGCTCAGCTTCTGCTCCAGTAGGCGCGCGCTCGGGC 635

6840 TrpArgProAlaPro- ArgAlaLysSerAlaAlaProSerArgThr- Arg*** 6857
634 CGATGTCAATGTGTGTCATGTCGAAGATCGATCCAGAAAGAAACGGCGCAGCAGGCA 575
6858 ProThrArgCys***Arg- ArgValSerGlyArgSerArgSer 6877
574 TCTGGCGCAGGCGCAGGTGCTACTCCCGCTCAGTAGAAGTCGGCAACGC- CCACCA 518
6872 Ser- AlaArgAlaArgTrpCysGlyTrpProGlyArgSerArgTrpProPro 6889
517 GGCCTCTTGGGACAGTACGAGC- AACCTGTGATGATGCTTGTGGACCGTG 464
6890 ThrArgThrProProGlySerSerMetProArgProGlyArgProArg- 6905
463 GAATCATCTGAGCGAGTAGAGAGCGTGCAGCGTGCCTCCAGGGCGCCAA 404
6906 ArgThrSerProValTrpArg- GlyArgCysTrpThrArgAlaSer 6921
403 ATTGGCAAGCTGCGCTCCACGCTTGACATTTCCAGCCCTTGGCCTTCTCTCT 344
6922 ThrGlySerAsnArgPro- ThrProAlaTrpSerProGly 6934
343 TCTTGGCCACTCCGACAGCGAGTGCACAGCTCGACCGTAGATGAGCTTGAACCTCG 284
6935 GlyTrpPro- ArgTrpAlaTrpProArg 6943
283 CCAGGT- CGA 275
6944 ProValSerIleGlyAlaSerSerAspCysProProValSerValTyrcysThrLeuArg 6963
274 TGTAAATCAGCCATCATATCGACATTTCTCCAGCTGCCACCCAGGT- 227
6964 CysSerArgProThrProArgProGlyProArgCysTrpThrPro***CysArgSerArg 6983
226 -----CAACCCAGATGAGGT 212
6984 ArgProProSerTrpAlaSerArgSerSerAlaTrpSerCysThrSerThrArgSerPro 7003
211 TCAGCGCTCGGCGCAGCGCGCAGCAAGAGCAGCGCTGGGCGCTAGAGCTTCCA 152
7004 SerThrAlaArg- HisArgArgArgArgSerArgSerArgSerIlePro 7020
151 GCGAGCAG- CGTATCATCGCCTTCTTGC 122
7021 AlaSerProThrProCysThrProTrpSerArgThrGlyArgAspAlaArgTrpArgAla 7040
121 TCCCGAACACATATGGCGCAGACCGTCAAGTCACTTCATGCTGCTGAGCTTGAGCT 62
7041 CysAlaAlaCysAlaThrArgTrpSerSerArgProArgArgThrProThrSerArg 7060
61 TCTCCAGGAGAGTCTTCTTGTGTAGTCTCGCAGCGCGCGCT 17
7061 AlaAlaThrArgThrArgThrTrpCysArgSerArgValCysArg 7075

RESULT 10
ABB98398
ID ABB98398 standard; protein; 19938 AA.
XX
AC ABB98398;
XX
DT 05-MAR-2003 (first entry)
XX
DE Streptomyces viridochromogenes Avi gene cluster polypeptide frame 1.
XX
XX Avilamycin; antibacterial; virucide; protozoacide; fungicide; infection;
XX medicine; Staphylococcus aureus; biosynthetic gene cluster; enzyme.
XX Streptomyces viridochromogenes.
XX
XX WO200268436-A1.
XX
PD 06-SEP-2002.
XX

PF 24-AUG-2001; 2001WO-EP009815.
 XX 25-FEB-2001; 2001DE-01009166.
 XX (COMB-) COMBINATURE BIOPHARM AG.
 PA Weitnauer G, Muehlenweg A, Trefzer A, Bechthold A;
 PI WPI; 2003-018650/01.
 XX N-PSDB; ABZ37515.
 DR New avilamycin derivatives, useful for treatment of infections, and
 XX nucleic acid encoding avilamycin synthesis enzymes.
 PT Example 1; Page 68-301; 319pp; German.
 XX
 XX The invention relates to avilamycin derivatives (I) with antibacterial,
 CC virucide, protozoacide and fungicide activity. (I) are useful for
 CC treatment of infections (bacterial, viral, protozoal or fungal), in human
 CC or veterinary medicine, particularly where caused by *Staphylococcus*
 CC aureus. (I) are more hydrophilic than known avilamycins. The present
 CC sequence is that of an avilamycin synthesis enzyme from the *Streptomyces*
 CC viridochromogenes Avilamycin A biosynthetic gene cluster (ABZ37515-
 CC ABZ37516)
 XX
 SQ Sequence 19938 AA;

Alignment Scores:
 Pred. No.: 6,52e-23 Length: 19938
 Score: 435.50 Matches: 230
 Percent Similarity: 33.7% Conservative: 49
 Best Local Similarity: 27.8% Mismatches: 201
 Query Match: 12.0% Indels: 347
 DB: 6 Gaps: 53

US-10-620-914-44 (1-1947) x ABB98398 (1-19938)

QY 23 GGCTCGGAGCTACA-----CCAGAAGA-----ACTTCTCCCTGGAGAAGCTCA 67
 DB 8581 GlycluargarghHisPro***ProarghHisTrpThrglyProargargser 8600
 QY 68 AGCTCAGCAGCATGAAGGATGACCTGACCGCTTCTGCGCCCATATGTGTTCGGCAGCAAGA 127
 DB 8601 GlySerAsp-----GlyThrAlaSerArg 8608
 QY 128 AGGCGGATGATCAGCTGCTGCGCTGGAGAGCTTCTACGGGCGCCAGG-----CCG 178
 DB 8609 Thr-----ArgAlaArgPheSerProArgThrArgSerPro 8620
 QY 179 CTG-----CCTTTGCTGCCCGCTGGCGAGGCTCGA 211
 DB 8621 ValSerArgArgAlaProAenGlySerSerProAlaSerProThrArgAlaCysAlaPro 8640
 QY 212 ACCTCATCTGGTTGACCTGGGTGGTGGCACTGGGAGAAATGTCGATATGATGCTGATT 271
 DB 8641 ThrThrThrGly----- 8644
 QY 272 ACATCGACCTGGCGAAGTCAAGT---CCATCTACGTGGTGGACCTGTGCCACTCGCTGT 328
 DB 8645 ThrSerProSerArgAlaMetSerValProSer-----SerThrAlaCys 8659
 QY 329 GCGAGGTGGCCAGAAGCAGGCGAAGGCCAAGGCGCTGMAA-----ATGTCAGGCTCG 382
 DB 8660 ThrThrTrp---ArgSerArgThrGlyProarghGlyThrCysCysThrAlaArgAsn 8678
 QY 383 TGGAGCGCAGCTTGGCAATTTTCGCCCTCGAGGCGACCGCGACGCTCATCACCTTCT 442
 DB 8679 TrpArgSerSer-----ProProArgSerTrpThrAlaArgSerCysProPro 8694
 QY 443 CTTACTCGCTCA-----CGATGATTCACCGTTCACACAG 478
 DB 8695 ProThrArgSerSerSerArgSerArgThrGlyProarghCysHisGlyThrValThr 8714

QY 479 TCATCCACCAGGCTT---GCTCGTACCTGTCCCAAGACGCGCTGTGGGGCTTGCCTGACT 535
 DB 8715 AlaSerThrSerAlaArgAlaArgSerAlaIle***AlaSerAlaTrpThrProAsp 8734
 QY 536 TCTACGTGAGCGGCAAGTACGACCTGCCCTGCGCCAGATGCCCTGGTCGGCGGTTCT 595
 DB 8735 ArgThrThrAspAlaTrpArg-----Arg 8742
 QY 596 TCTGGCGATCGATCTTCGACATCGACAACATTCGACATCGGCCCGGAGCGCGCGCT--- 652
 DB 8743 PheArgAspArgThrCysSerProThrThrProArgTrpArgArgCysAlaArgProAsp 8762
 QY 653 ACCTGAGCAGAGAGCTCGAGCGCG-----TGTGGGAGCAGA 688
 DB 8763 ProTrpPheArgSerArgSerAlaArgAlaThrTrpCysThrThrSerGlySerSer 8782
 QY 689 ACACCCAGGTTGATCCCTACGTGCGGTGGC----- 721
 DB 8783 ThrValArgAlaArgThrProThrAlaValGlyGlyGlyProSerSerSerThrPro 8802
 QY 721 ----- 721
 DB 8803 ThrArgProHisArgArgArgProGluArgCysAlaAlaGluArgGlnAlaAspIleGlu 8822
 QY 722 -----TGGCGCGCCCT----- 733
 DB 8823 ArgGlnValArgGluSerPheGly***GlyLeuCysSerProProArgThrGlyArgAsp 8842
 QY 734 ACTAGTGTGGATTGGCGCGCTCCAGCGCTTGGCCAGCGCC---TGCACGAGAGCGCG 790
 DB 8843 IleThrCysAlaTrpCysProTrpProGlyArgCysAlaProArgAspThrArgSerGly 8862
 QY 791 TGG-----AGCGCGCGCCCATGTTCCCGCCACCTTCTCTGTACA 829
 DB 8863 TrpArgAlaArgArgArgArgAenAlaGlySerPro-----ArgProAla***Cys--- 8879
 QY 830 CGCAGTGTGGGAGGACCCCGAGCGGATATGGAGG-----TCATGAGATCAACCCCA 883
 DB 8880 ArgSerArgCysSerThrArgArgThr***TrpLysAlaProGlyTrpArgSer----- 8897
 QY 884 AGGACAGCTGTGTCAGCTGACTAGCGGGCTGCAATGCCCTGAACCTCTGTGTC--- 940
 DB 8898 ---MetCysArgPro-----CysThr 8903
 QY 941 -----AGGGGCGCGCCAGGTGG---TGTCTGTGGACTGCAACCCCGCAGTCGCGCG 991
 DB 8904 HisArgArgSerProValArgGlyArgCysArgCysThrArgSerProGlySerArg--- 8922
 QY 992 TTCTGGAGCTGAAGAAGGTGGCCATTTCAGCAGCTGGAGTTTGAGGACGTGTGGCAGTGT 1051
 DB 8922 ----- 8922
 QY 1052 TCGCGAGGCGGTGCACCCGCGCATTTGAGGAGCTGTACGAGAAGAGAGCTGGCGCCTTCC 1111
 DB 8923 -----CysSerArgSerThrThrSer----- 8929
 QY 1112 TGTGCAAAACAGCCACA-----ACTTCTGTGCCAAGCGCTCTGTACTTCCAGCAGC 1165
 DB 8930 -----ThrProAlaThrCysAlaThrSerGlyArgSerArgSerThrArgCysSerAla 8947
 QY 1166 GCCTGTACTACGAGCGGCATGGGCAAGCTGTGCTGG-----TGCTGAGT 1213
 DB 8948 AlaMetThrThrArg***AlaSerAlaThrThrGlyGlyProThrTrpTrpCysThrThr 8967
 QY 1214 GCCTGGCGGTGTGCTGGGACTGGCA----- 1240
 DB 8968 SerTrpProSerArgAlaHisTrpSerProArgSerGlyAlaCysArgAlaSerMetCys 8987
 QY 1241 AGACCGTCAAGCGCTCGCCAAACGCGCCCAAA-----TGG--- 1276
 DB 8988 ArgProGlySerSerAlaProTrpArgProSerArgGlySerThrTrpSerArgThr 9007
 QY 1277 -----AGGAGCAGCGCTGTGTGGGACAGCAACATGCTCATCCATTCG 1321

[illegible]

```
QY 1585 GCAGGTAGTGGGCGAGTGTCTCGCGCAGGAACCTTCGGGTGAGGCGAGTTGTAGTACA--- 1529
Db 17562 ArgGlyArg-----ArgGlnAlaGlyCys-----AlaCysCysSerArgGly 17575
QY 1529 ----- 1529
Db 17576 GlyProHisGlyAspGlyArgTrpArgProTrpArgArgSerAlaHis***ProSerArg 17595
QY 1528 ---AGTAGTCTCTGTCGCGACGTGCGAGTCTCCGCCACGCGTCCATGTCGCGCGCA 1472
Db 17596 ProThrThrValCysHisArgCysSerProArgProArgSerThrCysCysAlaAla 17615
QY 1471 TGTAGTCTCAATGG-----GGATGCCGTCCGCTTCATCGCGCGT 1430
Db 17616 ArgAlaProArgTrpProProTrpSerThrSerSerCysAlaProPro-----SerThr 17633
QY 1429 ACTGCTGCGCGCAGCGCGCGCGGACACACAGCAGCGCTTGTGGAAGAGCACCA--- 1373
Db 17634 ThrAlaCysSerProSerSerArg-----AlaAlaProProThrProArgProArg 17651
QY 1372 -----GGCTCAGCAACTTCACGAACAGCCACA 1346
Db 17652 GlySerCysArgThrArgTrpProSerGlySerGlyProArgArgSerThrThrArgThr 17671
QY 1345 CCAGCGGCTTGGCGCGCTTCTCACGAAGTGGATGAGCATGTTGCTGCTCCACAGACGCG 1286
Db 17672 AlaGlySerHisCysAlaAlaThrArgProSer---AlaProAlaSerProSerProSer 17690
QY 1285 GCTGCTCTCCATGTGGCGCGTGGCGAGGCGCTTGACGGTCTGGCCAGTCCACGCA 1226
Db 17691 AlaGlyArgSerSerAlaThrArgTrpGlySerAlaAlaArgSerArgProSerProHis 17710
QY 1225 CCAGG-----CCAGGCACTGCAGCACCAGC-----ACAGCTTGC 1190
Db 17711 SerArgThr***ProProArgGlyArgAlaThrSerTrpSerValArgProAlaCys 17730
QY 1189 -----CCATCGCGCTCTGGTAGT-----ACA 1169
Db 17731 ***SerThrGlyProGlyArgProArgSerAlaValCysCysAla***HisArgCysThr 17750
QY 1168 GGCGGTGCTGGAAGTACCAGAGCGCTTGGACCAGAAAGTTGTGGCTGTTTGGCAGGGA 1109
Db 17751 AlaProThrGlyArgThrArgSer-----ProThrGly 17761
QY 1108 AGGCGCGCAGCTTCTCTGACAGTCTCCTCAATGCGCGGTGCACGCGCTCCCGNACA 1049
Db 17762 ArgGlyGlyAlaAlaSerArgSerProSerThrArgCysGlyAlaSerArgArgPro 17781
QY 1048 GCTGCC-----ACAGTCTCTCAAACT----- 1028
Db 17782 GlyAlaAlaArgTrpAlaCysSerSerTrpThrArgProThrThr***ArgIleProGly 17801
QY 1027 CCAGCTGCTGAATGGCCACCTTCTTCAGCTCCAGAACGCGCGTACGTCGCGGGTTGCAGT 968
Db 17802 ProAlaAlaProTrpProSerProSerGlyArgSerThrAlaThrAla---CysCysSer 17820
QY 967 CCACCGACACACTGCGCGCGCGCGCGTGCACCAGCAGTTCAGGGCATTCGACGCGCGCG 908
Db 17821 ***ProValArgProTrpArgThrAlaSerArgSerSerAlaAlaTrpCys----- 17837
QY 907 TAGTCAGGTCAGCACCGTGTCTTGGGGTTGATCTCCATCACCTCCATCCGCGTCCG 848
Db 17838 -----GlySerSerProSerTrp-----Arg 17845
QY 847 GGTCTCCACGACTGGGTGTACAGGAAGTGGCGGGAACATGGCGCGCTCCACGC 788
Db 17846 ArgProSerThrSer-----ThrThrAlaTrp-----Arg 17855
QY 787 GCTCTCTGTCAGGCGGTGGCCAAAGCTGGCGAGCGCGCCCATCCACGCTAGTAGGGG 728
Db 17856 AlaProArg-----ArgSerAlaArgSerHis 17865
```

```
QY 727 CGC-----GCAGCCACGCGACCTAGGGGATCGAACCCCTGGGTGTCTGTCTCCACA 677
Db 17866 ArgSerIleCysAlaAlaThrSerArgThrSerSerProAsnSerArgArgSerThr 17885
QY 676 CGCGCTCCAGCTTCTGCTCAGGTAGCGCGCGCTCGGGCCGATGTCAATGTTGCGA 617
Db 17886 ArgThrSerGlyArgSerProAlaArgArg----- 17895
QY 616 TGTGCAAGATCGATCGCCAGAAACGCGCGGACCCAGGGCATCTCGCGCAGGGCAGGT 557
Db 17896 -----ThrArgArgProThrAlaArgProCysAlaProAlaThrSerTrp 17910
QY 556 CGTACTTTCGCTCAGTAGAAGTCGCAACGCCACCGCGCTCTTGGGACAGGTACG 497
Db 17911 Arg---CysAlaGlyArgArgThrArgValArgSerGlyProArgArgCysThrGlyCys 17929
QY 496 AGCAACGCTGTCGATGAGCTTGTGGAACGGTGAATCATCGTGAGCGAGTAGAGAAG 437
Db 17930 AlaArg----- 17931
QY 436 TGATGAGCGTTCGCGGTGCCTCAGGGGCGCAATTGGCAAGCGTCGCGCTCCACGACCT 377
Db 17932 -----SerLeuArgArgProProArgThr 17939
QY 376 GGACATCTTCAGCCCTTGGCCTTCGCTTCTCT-----TGCCA----- 335
Db 17940 Gly***ArgSer-----TrpCysSerProArgSerGlyThrTrpTrpGlnTrpTrpAla 17957
QY 334 -----CCTCGCACGCGAGTGCACAGCTCGACACAGT---AGATGGACTTGAACCTCG 284
Db 17958 ArgArgSerArgThrArgProGlyGlnGlyArgGlyGlnAlaArgArgThrSerArgGly 17977
QY 283 CCAGGTGCGATGTAATCAGCCATCATATCGACATCTCTCCCGAGTCCACCCAGGTCGCA 224
Db 17978 ProLysArg---GlyGluProGlnArgMetArgGlyAlaArg---ArgTrpLysGlyGln 17995
QY 223 CCAGATGAGGTTTCGACGCGCTCG-----CCAGCGCGGCGAGCAAGAGCAG--- 179
Db 17996 ProArgArgThrArgArgAlaArgArgArgSerProProTrpGlyThrHisArgArg 18015
QY 178 -----CGGCTGGCGCGCTAGAGCTCTCCAGCGGAGCAGCGTGATCAT 134
Db 18016 ArgSerGlyArgGlyArgAlaGlyCysArgAspArgGlyProGlyValCysSerGlyArg 18035
QY 133 CGCCTCTTCTGCTGCCGA-----ACCATATATGCGCGAGAACGG 95
Db 18036 SerArgAlaAlaCysArgProProAlaAspSerSerSerSerThrThrSerProArg 18055
QY 94 TCAGGTATCCTTCATGCTGCTGAGCTTTCAGGGAGAGTCTTCTTCTTGTGTGT 35
Db 18056 ProValThrProTrpCys-----SerArgArgSerArgGlnAlaGluTrpAla 18071
QY 34 AGC-----TCGACGCGCGCGCTCAGCCCGGACCCCA 2
Db 18072 SerThrCysArgProProTrp***SerSerAlaSerArgSerSerArgPro 18090
RESULT 12
ABP76682
ID ABP76682 standard; protein; 19938 AA.
XX AC ABP76682;
XX DT 26-FEB-2003 (first entry)
XX DX Streptomyces viridochromogenes Avi gene cluster polypeptide frame 6.
XX DE Streptomyces viridochromogenes
XX KW Avilamycin; antibacterial; virucide; protozoacide; fungicide; infection;
XX KW medicine; Staphylococcus aureus; biosynthetic gene cluster; enzyme.
XX OS Streptomyces viridochromogenes.
XX PN WO200268436-A1.
XX
```


QY	517	-----GGCCGCTTTGGACAGGTACGACGAGCTGCTGCATGACGT	476
Db	11120	CysLeuAlaLeuGlyGlyAlaAlaLeuArgAlaProAlaValArgProGlyArg-----	11137
QY	475	TGTGGAACGGTGAATCATCTGAGCGAGTAGAGAGAGGTGATGAGCGTCCGGTGC---	419
Db	11138	-----ArgThrArg***ArgTrpSerAlaAlaAsnCysArg	11149
QY	418	-----CCTCAGGGGGCGCAATWGCAAGGTCGGCTCCA---CGACCTGGA	374
Db	11150	TrpArgSerGlyProAsnArgGlyArgAlaGlyArgAlaProArgProGly	11169
QY	373	CATTCTCCAGCCCTTGGCTTCGCTCCT-----TCTTCTGGCCACCTCGC	329
Db	11170	-----***ProGlyProGluProArgValArgSerArgAlaProSerProProArg	11186
QY	328	ACAGCGAGTGGCAGGT-----CGACCACGT	302
Db	11187	ArgArgArgGlyAlaGlyAlaIleProGluProProGlyIleArgCysAlaValArg	11206
QY	301	AGATGACCTTGAACCTCGCCAGCTCGATGTAAT-----CAGCATCATATCGACATTCT	248
Db	11207	Arg-----ArgProGlyArgCysSerAspCysArgProCysAlaGlyArgArg	11222
QY	247	CCCAGTGCACCACCCAGGTCAACCCAGATGAGTTCCGAGCGCTCGGCCAGCGCGGCGAG	188
Db	11223	***ArgGlyHisGlyAlaMetAlaPro-----ProArgProGlyThrAla	11237
QY	187	CAAAGCGAGCGGCTGGGGCCCGTAGAAGCTCTCCAGCGAGCAGCGGTGATCATCGC---	131
Db	11238	Pro***GlyArgAlaArgArgTtpAlaArgProGlyArgProArgThrArgArgArg	11257
QY	130	-----CCT	128
Db	11258	AlaAlaProValProArgGlyAlaAlaGlyProValProGlyProSerProAlaSerPro	11277
QY	127	TCTTGTCTCGCGAACCATATATGGCGCAGAACGGTTCAGGTTCATCTTCATGCTGTGAGCT	68
Db	11278	SerCysAlaArg-----GlyArgGlyArgAspAlaHis-----	11288
QY	67	TGAGCTTCTCCAGGAGAGTCTTCTTGTGTAGTCTCGCAGGCCGCGC	20
Db	11289	-----ArgProArgArgGlySerSerSerSerArgArgSerAlaGly	11302
RESULT 13			
ID	ABP76678		
AC	ABP76678 standard; protein; 19938 AA.		
CC	ABP76678;		
XX	26-FEB-2003 (first entry)		
DE	Streptomyces viridochromogenes Avi gene cluster polypeptide frame 2.		
KW	Avilamycin; antibacterial; virucide; protozoacide; fungicide; infection;		
KW	medicine; Staphylococcus aureus; biosynthetic gene cluster; enzyme.		
XX	Streptomyces viridochromogenes.		
XX	WO200268436-A1.		
XX	06-SEP-2002.		
XX	24-AUG-2001; 2001WO-EP009815.		
XX	25-FEB-2001; 2001DE-01009166.		
XX	(COMB-) COMBINATURE BIOPHARM AG.		
XX	Weitnauer G, Muehlenweg A, Trefzer A, Bechthold A;		
XX	WPI; 2003-018650/01.		
DR			
N-PSDB; ABZ37515.			
New avilamycin derivatives, useful for treatment of infections, and			
nucleic acid encoding avilamycin synthesis enzymes.			
Example 1; Page 68-301; 319pp; German.			
The invention relates to avilamycin derivatives (I) with antibacterial,			
virucide, protozoacide and fungicide activity. (I) are useful for			
treatment of infections (bacterial, viral, protozoal or fungal), in human			
or veterinary medicine, particularly where caused by Staphylococcus			
aureus. (I) are more hydrophilic than known avilamycins. The present			
sequence is that of an avilamycin synthesis enzyme from the Streptomyces			
viridochromogenes Avilamycin A biosynthetic gene cluster (ABZ37515-			
ABZ37516)			
Sequence 19938 AA;			
Alignment Scores:			
Pred. No.:	3,84e-20	Length:	19938
Score:	398.50	Matches:	225
Percent Similarity:	33.5%	Conservative:	46
Best Local Similarity:	27.8%	Mismatches:	265
Query Match:	11.0%	Indels:	273
DB:	6	Gaps:	46
US-10-620-914-44 (1-1947) x ABP76678 (1-19938)			
QY	23	GGCTCGGAGTACACCAAGAGAACTTCTCCTCGG-----AGAAGCTCAAGC	70
Db	17449	GlyCysArgThrSerProArgAlaGlyCysAlaTrpGlyArgTrpArgArgAlaGlySer	17468
QY	71	TCAGCAGCATGAAGGATGACCTGACCGTCTTCTGCGCATATGTGTTCTGCGGCAAGAAGG	130
Db	17469	GlyArgSer-----AlaArgCysSerThrProAlaArg--	17479
QY	131	GCATGATCAAGCTGCTGCGCTGGAGAGCTTACGGGCCCCAGCGCTGCTTGCCTG	190
Db	17480	-----ThrGlyCysAlaSerSerAlaGlySerGlySerArgPro-----	17492
QY	191	CCGCGC-----TGGCGGAGCGCTCGAACCTCATCTGGTTGACCTGGTGGTGCGACTG	244
Db	17493	ProThrArgArgTrpProProSerGlySerProArgLeu-----Tyr	17507
QY	245	GGCAGAAATGCGATATGATGGCTGATTACATCCAGCTGCGCAAGTTCACATCTCATCG	304
Db	17508	GlyThrAlaSer-----ProTyrThrSerThrTrpThrGlyProSerProGlyPro	17524
QY	305	-----TGTCGACCTGTGCCACTGCTGTGCGAGGTGGCCAGAGAGCGGAAGG---	355
Db	17525	ProHisTrpSer-----SerArgCysThrCysTrpTrpArgArgAlaProArgHis	17541
QY	356	-----CCAAGGGCTGAAGAATGTCAGGTGC-----	382
Db	17542	AlaGlyProSerArgArgAlaGlyArg***ArgArgSerSerAlaArgCysTrpProArg	17561
QY	382	-----	382
Db	17562	ArgGlyArgGlnAlaGlyCysAlaCysCysSerArgGlyProHisGlyAspGly	17581
QY	383	-----TGGAGCGCGAGCTTGCCTGCAATTTGCGCCCCCTGAGGCGACCG-----	424
Db	17582	ArgTrpArgProTrpArgArgSerAlaHis***ProSerArgProThrThrValCys	17601
QY	425	CGAGCTCATCAGCTTCTCTACTCGCTCAGTATGATTCACCGTTCCAGACCTCATCG	484
Db	17602	HisArgCysSerProArgPro-----ArgSerThrCysCysAla	17614
QY	485	ACCAGGCTTGTCTGCTACCTGTCCTCAAGACGCGCTGTTGGGGCTGCTTCTTACGTGA	544
Db	17615	AlaArgAlaProArgTrpProProTrpSerThrSerSerCysAlaProProSerThrThr	17634
QY	545	GCGCAAGTACGACCTGC-----CCCTGCGCCAGATGC	577

Db 17635 AlaCysSerProSerArgAlaAlaProProThrProArgProArgGlySerCys 17654
QY 578 CTGTGTCGCGCCGTTCT---TCTGGCATCGATCTTCGACATCGACCAACATTGACATCG 634
Db 17655 ArgThrArgTrpProSerGlySerGlyProArgSerThrThrArgThrAlaGlySer 17674
QY 635 -----GCCCGAGCGCGCGCTTACC 655
Db 17675 HisCysAlaAlaThrArgProSerAlaProAlaSerProSerAlaGlyArgSer 17694
QY 656 TGGAGCAGAGCTGGAGCGGTGTGGAGCAGACACCCAGGTTTCGATCCCTTACGTGC 715
Db 17695 SerAlaAlaThrArgTrpGlySerAlaAlaArgSerArgProHisSerArgThr--- 17713
QY 716 CGTGGCTGCGCGCCCTTACTAGCTGTGATGGCGCGCTGCGCAGCGTGTGGCAGCGCC 775
Db 17714 -----**ProProArg-----GlyArgAlaThrSerTrpTrpSerValArgPro 17728
QY 776 TGCACGAGGAGCGGTGGAGCGCGCCCA-----TGTTCGCGCCACCTTCC 823
Db 17729 AlaCys**SerThrGlyProGlyArgProArgSerAlaValCysCysAla**HisArg 17748
QY 824 TGTACAGCAGCTGTGGAGGACCCCGAGCGGATATGGAGGTGATGG-----AGA 874
Db 17749 CysThrAlaProThrGlyArgThrArgSerProThrGlyArgGlyAlaAlaSerArg 17768
QY 875 TCAACCCAGGACGAGTGTGCTGACCTGACTAGCGCGCTGCAATGCCCTGAACCTGC 934
Db 17769 SerProSerThrArgCys-----GlyAlaSerArgPro----- 17781
QY 935 TGTGTCAGGGCGCGCCAGTGTGT-----CGGTGACTGCAACCCCGCGCAGTCGG 988
Db 17782 -----GlyAlaAlaArgTrpAlaCysSerSerTrpThrArgProThr***Arg 17798
QY 989 CGCTTCTGGAGCTGAAGAGTGGCCATTGAGCAGCTGGAGTTGAGGAGTGTGCAGC 1048
Db 17799 IleProGlyProAlaAlaProTrpProSerProSer-----GlyArg 17812
QY 1049 TGTTCGCGAGCGGTGCACCCCGCATTTGAGGAGCTGTACGAGAGAGAGTGGCGCCCT 1108
Db 17813 SerThrAlaThrAlaCysCys-----Ser***ProValArgProTrpArgThr 17828
QY 1109 TCTGTGCGAAACAGCCCACTTCTGTGTCAGAGCGCC-----TCTGTT 1153
Db 17829 AlaSerArgSerSerAlaAlaTrpCysGlySerSerProSerTrpArgProSer 17848
QY 1154 ACTTCCAGCAGCGCC-----TGTACTACGAGCGCGCTGCGCAAGCTGTGCTGGTGC 1207
Db 17849 ThrSerThrThrAlaTrpArgAlaProArgArgSerAlaArgSerHisArgSerIle 17868
QY 1208 TGCAGTGTGCGCGTGTGCTGGAGCTGGGCAAGACCGTCAAGC-----GCC 1255
Db 17869 CysAlaAla-----ThrSerArgThrSerSerProAsnSerArgArg 17882
QY 1256 TGCACCAACCGCCCACTGGAGGAGCAGCGCC----- 1288
Db 17883 SerSerThrArgThrSerGlyArgSerProAlaArgArgThrArgProThrAlaArg 17902
QY 1289 GTCTGTGGGACAGCA-----TGCTCATCC-----ACTTCGTGAAGACGGCG 1333
Db 17903 ProCysAlaProAlaThrSerTrpArgCysAlaGlyArgArgThrValArgSerGly 17922
QY 1334 CCAAGCGCTGTGCTGTGCTGTGCTCAAGT-----TCGTGA 1369
Db 17923 ProArgArgCysThrGlyCysAlaArgSerLeuArgArgProProArgThrGly***Arg 17942
QY 1370 GCCTGTGTC-----TCTCAACA 1387
Db 17943 SerTrpCysSerProArgSerGlyThrTrpTrpGlnTrpTrpAlaArgSerArgThr 17962
QY 1388 AGCCCGTGTGCTGCGCGCGCGGTGCGCGCAAGC----- 1426

Db 17963 ArgProGlyGlnGlyArgLysGlnAlaArgThrSerArgGlyProLysArgGlyGlu 17982
QY 1427 -----AGTACGCGCTGATCAAGCGCGCGCATCCCATTTGAGAACTACATCGCGCGCA 1480
Db 17983 ProGlnArgMetArgGlyAlaArgArgTrpLysGlyGlnProArgArgThrArgAla 18002
QY 1481 -----CCATGACCGCGTGGCGGAGAACTCGCAGC----- 1510
Db 18003 ArgArgArgSerProProTrpGlyThrHisArgArgArgSerGlyArgGlyArgAla 18022
QY 1511 -----TGGCAAGCAGCACTACTTCTACTACACTGCC 1543
Db 18023 GlyCysArgAspArgGlyProGlyValCysSerGlyArgSerArgAlaAlaCysArgPro 18042
QY 1544 TCACCGCGCAAGTTCTTCGCGCAACTGCCCACTGCCACCTACC-----TGGCGC 1588
Db 18043 ProAlaAspSerSerSerSerThrThrSerProArgProValThrProTrpCysSer 18062
QY 1589 AGCGCGCTTCCGACCTCAAGAGTGGCGTGGTGACAACTGACCTGACCTCCACCA--- 1645
Db 18063 ArgArg-----SerArgGlnAlaGluTrpAlaSerThrCysArgProProPro 18078
QY 1646 -----ACTTCTCATGGAGGAGCTCA-----AAGCGCGCACCTACA 1681
Db 18079 Trp***SerSerAlaSerArgSerSerArgProSerAsnThrArgArgSerProVal 18099
QY 1682 CCAAGGTGATTTCTGATGGACCACTGCGACTGCGTGG----- 1717
Db 18099 Pro-----ThrAlaTrpAlaArgSerAlaArgSerAlaCys 18110
QY 1718 -----ATATGCGCGTGGCAACGAGCTGCGCAGAGTGCCTGGCAAGAGTTGCGCGCGCG 1774
Db 18111 ThrAlaCysSerAlaAlaArgLysGlyTrpThrAsnAlaTrpCysGlyCysTrpArgThrSer 18130
QY 1775 GCATCGTCTATCGCGCTCGGCTCCCTCAGCGCGCTCAGCGCGCTGATCCAGAGG 1834
Db 18131 ProAlaCysSerThrLeuThrProAlaAlaAlaArg-----SerProSerArgHis 18147
QY 1835 CGGCGCTTCAGCGTGCCTGCTGCTGCGCGCGCCACTCAGGCGCTACATGCGCGCTCAACA 1894
Db 18148 Arg-----ThrProSerThr 18152
QY 1895 TGTACAGCTCTTCTACATGCGCGCGCC 1921
Db 18153 ---SerArgThrSerProTrpProAla 18160
RESULT 14
AAW87503
ID AAW87503 standard; protein; 1212 AA.
XX AAW87503;
AC AAW87503;
DT 23-FEB-1999 (first entry)
XX
DE Human N-methyl-D-aspartate receptor subunit encoded by clone NMDA22.
XX
KW Human; N-methyl-D-aspartate receptor; NMDAR2C;
KW NMDA-activated cation-selective ion channel; glutamate receptor.
XX Homo sapiens.
XX US5849895-A.
XX 15-DEC-1998.
PD
XX 20-APR-1994; 94US-00231193.
PF
XX 20-APR-1993; 93US-00052449.
PR 20-APR-1994; 94US-00231193.
XX
PA (SIBI-) SIBIA NEUROSCIENCES INC.
XX
PI Lu C, Daggett LP;

Db 692 ArgArgThrProThrGlySerGlyProThrTrpArgSerCysSerSerTrpGlyThrGlu 711
Qy 1157 -----TCCAGCAGCGGCTGTACTACCAAG 1180
Db 712 ArgHisArgAenTrpArgGlnCysGlySerGlnGlySerAlaArgMetArgArgThrArg 731
Qy 1181 GCGGCA-----TGGCAAGCTGTCTGGGTCTGCAGTGTCTGGCG 1222
Db 732 AlaAlaSerTrpThrSerThrTrpThrTrpGluAlaSerSerThrCysCys---TrpTrpPro 750
Qy 1223 TGTGCTGGGAC---TGGCAAGACCGCTCAAGCGCC-----1255
Db 751 TrpGlyTrpProCysTrpSerSerProGlySerThrTrpSerThrGlySerCysAlaThr 770
Qy 1256 ---TGGCAACGCGCCCAATAGGAGGAGGAGCGCCGTC---TGTGGCAGCAGCAATGC 1309
Db 771 ArgCysProThrHisProSerTrpThrSerCysTrpLeuSerAlaGlyAlaSerThrAla 790
Qy 1310 TCATCCACTTCGTGAAGACGGGCCA-----1336
Db 791 AlaSerAlaGlyCysArgAlaSerProAlaHisArgGlyArgProAlaArgThrSerArg 810
Qy 1337 -----AGCGCTGGGTGGCTGTCTGCTCA-----1360
Db 811 ProAlaArgProArgProAlaCysSerArgPheCysArgGlnProAlaThrTrpProArg 830
Qy 1361 ---AGTTCGTAGCCTGTCTCTTCAACAGGCGGTC---TGTGCTCGCGCGCGG 1414
Db 831 ArgAlaAlaThrProTrpThrAlaProLeuAlaProSerArgGlyValAlaAlaAla 850
Qy 1415 -----TGGCGGCAAGCAGTACGCGCTGATCAAGCGGACGGCA 1453
Db 851 ValArgProHisArgProProAlaArgProArgGlyLeuAlaProAlaHisAlaCysPro 870
Qy 1454 TCCCAATGTAGAATACTACATCGCGGCACCATGGACGCGTGGCGAGAACT-----1504
Db 871 ProProThrArgProGlnSerArgAlaProArgAlaGlyAspArgGlnThrGlyValAla 890
Qy 1505 ---CGCAGTGGCGAAGC---AGAACTACTTCTACTACAACCTGCCTCACCAGGATCC 1558
Db 891 ArgArgLeuCysAlaGlyLeuAlaGlySerProArgAlaAlaProArgArgGlyArgPro 910
Qy 1559 TGGCGGCAACTGCCCACTACTTGC-----1585
Db 911 CysProThrSer-----ProGluCysArgAlaAlaGlnProGlyArgArgGlyArg 928
Qy 1586 -----GCGAGCGGCTTCCGCCACCTCAAGAGTGGCG 1618
Db 929 CysGlyProGlyThrAlaGlyGlyThrSerArgProSerGlyProCysArgProArg 948
Qy 1619 TGTGGACAACTGACCG-----TCTCCACCACTTCTCATGGAGAGCTCAAGCGC 1672
Db 949 AlaValThrAlaProPheLeuGluProThrAspProAlaAlaProSerSerArgSer 968
Qy 1673 GCACCTACACCAAGGTGATCTGATGACACAG-----TGG 1708
Db 969 SerArgSerProArg-----SerTrpArgThrCysArgCysSerValArgSerSerTrp 986
Qy 1709 ACTGGC-----TGATA 1720
Db 987 ProGlyGlyArgProCysThrArgProGlyProGlyAlaArgAlaArgValThrLeuPro 1006
Qy 1721 TGCCCGTGGCGCAGAGCTGGCGAGTGCCTGGCCAGCAGGTTGGCGCGGCGGATCG 1780
Db 1007 CysProAlaProTrpProArgProSerLeuGlyProAlaArgCysProLeuGlyAlaPro 1026
Qy 1781 TCATCTGGCGCTGGCTCTCCCTGAGC---CGCCTACGCGGAGCTGATCCAGAGCGG 1837
Db 1027 AlaProProAlaProAlaProThrAlaThrArgProAlaGlyAlaTrpArgArgSer 1046
Qy 1838 GCTTCGACGTCGTGATCGCGCGGCGGCACTCAG-----1873
Db 1047 ArgCysAlaCysArgSerThrGlyArgProAlaArgAlaSerArgGlnGlyProPro 1066

Qy 1874 -----GCTACATGGACCGCTCAACATGTACAGCTCTCTTCTACA 1912
Db 1067 ProGlySerThrAspSerThrSerAlaCysThrProThrProThr 1081
RESULT 15
AAB26239
ID AAB26239 standard; protein; 1232 AA.
XX AAB26239;
DT 23-FEB-2001 (first entry)
XX Human N-methyl-D-aspartate receptor subunit NMDAR1A #17.
XX Human; N-methyl-D-aspartate receptor; NMDA; NMDAR1A; ionotropic;
KW glutamate receptor; drug screening; animal model; disease diagnosis;
KW genetic screening.
XX Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 280 /label= unknown
FT /note= "encoded by TGA"
FT Misc-difference 304 /label= unknown
FT /note= "encoded by TGA"
FT Misc-difference 326 /label= unknown
FT /note= "encoded by TGA"
FT Misc-difference 347 /label= unknown
FT /note= "encoded by TGA"
FT Misc-difference 422 /label= unknown
FT /note= "encoded by TGA"
FT Misc-difference 438 /label= unknown
FT /note= "encoded by TGA"
FT Misc-difference 464 /label= unknown
FT /note= "encoded by TAG"
FT Misc-difference 477 /label= unknown
FT /note= "encoded by TAG"
FT Misc-difference 501.502 /label= unknown
FT /note= "encoded by TGATGA"
FT Misc-difference 590 /label= unknown
FT /note= "encoded by TGA"
FT Misc-difference 712 /label= unknown
FT /note= "encoded by TAG"
FT Misc-difference 744.745 /label= unknown
FT /note= "encoded by TGATGA"
FT Misc-difference 843 /label= unknown
FT /note= "encoded by TGA"
FT Misc-difference 848 /label= unknown
FT /note= "encoded by TAA"
FT Misc-difference 1009 /label= unknown
FT /note= "encoded by TGA"
FT Misc-difference 1152 /label= unknown
FT /note= "encoded by TAG"
FT Misc-difference 1209 /label= unknown
FT /note= "encoded by TGA"
FT

```

FT Misc-difference 1227
FT /label= UNKNOWN
FT /note= "encoded by TGA; the coding sequence of this
FT protein contains a number of in-frame stop codons which
FT are represented by xaa in the protein sequence"
XX
XX US6111091-A.
XX
XX 29-AUG-2000.
XX
XX 29-SEP-1997; 97US-00940086.
XX
XX 20-APR-1993; 93US-00052449.
XX
XX 20-APR-1994; 94US-00231193.
XX
XX (MERI ) MERCK & CO INC.
XX
XX Daggett LP, Lu C;
XX
XX WPI; 2000-578607/54.
XX
XX N-PSDB; AAA95032.
XX
XX Novel DNA fragment encoding human N-methyl-D-aspartate receptor subunit
XX for identifying mutations and for developing drugs against various
XX disease states.
XX
XX Example 3; Col 255-264; 205pp; English.
XX
XX The present sequence is a subunit (designated NMDARIA) of the human N-
XX methyl-D-aspartate (NMDA) receptor. This is an ionotropic glutamate
XX receptor which contains cation-specific ligand-gated ion channels. The
XX protein and its coding sequence can be used in disease diagnosis and in
XX research to identify other, similar proteins. They can also be used as
XX probes, for example in genetic screening, and in drug screening, as well
XX as enabling the production of animal disease models
XX
XX Sequence 1232 AA;
XX
XX Alignment Scores:
XX Pred. No.: 1.09e-19 Length: 1232
XX Score: 386.00 Matches: 227
XX Percent Similarity: 32.4% Conservative: 76
XX Best Local Similarity: 24.3% Mismatches: 280
XX Query Match: 10.7% Indels: 353
XX DB: 3 Gaps: 53
XX
XX US-10-620-914-44 (1-1947) x AAB26239 (1-1232)
XX
XX 11 GTCGTGAGCGCGCGCTGCGAGCTACACCAAGAGAACT----- 49
XX ||||| ||| :|||
XX 216 ValValThrGluSerTrpArgLeuSerLeuArgGlnLysValArgAspGlyValAlaIle 235
XX ||||| |||||
XX 50 -----TTCCTCTGAGAGAGCTCAAGCTCAGCATCAGCAGCATGAGATGACC 91
XX ||||| |||||
XX 236 LeuAlaLeuGlyAlaHisSerTyrTrpArgGlnHisGlyThrGlnLys----- 251
XX ||||| |||||
XX 92 TGACCGTTCTCGGCATATGTTGTCGCACAGAGAGCGGATGATCAGCTG----- 145
XX ||||| |||||
XX 252 -----GlyValCysGlnProArgProGlyThrAlaValPheThrLeuGlyPro 267
XX ||||| |||||
XX 146 -----CTCGCTCGAGAGCTTCTACGGC----- 169
XX ||||| |||||
XX 268 SerAlaLeuProGlyArgProSerThrGlyThrTyr***MetSerProGlyArgAlaGlu 287
XX ||||| |||||
XX 170 CCCAGCGCGTGCCTTCTGCTCCCGCCTTG-----CCGAGC 205
XX ||||| |||||
XX 288 ThrSerProSerAlaLeuValGlyThrTrpSerSerProProTrpTrp***SerProSer 307
XX ||||| |||||
XX 206 GCTCGAACCTCATCTGGTTGACCTGGTGTGGCTGCTGGGAGAAATCTCGATATGATGG 265
XX ||||| |||||
XX 308 ThrGlyThrAlaSerGlyArgTrpTrp-----GlyAlaGlySerMet----- 321
XX ||||| |||||
XX 266 CTGATTACATCGACCTGCGGAAGTTCAAGTCCA----- 298
XX ||||| |||||

```

Db 660 GlySerTrpMetProSerSerMetMetLeuLeuSerSerThrThrTrpGlnAlaArgThr 679
Qy 1070 CGCCATTGAGGAGCTGTACGAGAAGAGCTGGCGCCCT----- 1108
Db 680 ArgAla-----AlaSerTrpSerProLeuGlyLeuAlaArgSerLeu 693
Qy 1109 -----TCCTGTCGCAACAGCCCAACTTCTGTGTC- 1141
Db 694 LeuProLeuAlaThrAlaSerProCysArgArgThrProThrGlySerGlyPro***Thr 713
Qy 1142 -----AGCGCTCTGGTACT----- 1156
Db 714 TrpArgSerCysSerSerTrpGlyThrGluArgHisArgAsnTrpArgGlnCysGlySer 733
Qy 1157 -----TCCAGCAGCGCTGTACTACCAAGCGCGCA 1186
Db 734 GlnGlySerAlaArgMetArgArgThrArg*****AlaAlaSerTrpThrSerThr 753
Qy 1187 TGGCAAGCTGTGCTGGTGTGTCAGTGCCTGGCCGTGGTGGGAC-----TGGGCAAGA 1243
Db 754 TrpGluAlaSerSerThrCysCys---TrpTrpProTrpGlyTrpProCysTrpSerSer 772
Qy 1244 CCGTCAAGCGCC-----TCGCCAAGCGCCCAATGG 1276
Db 773 ProGlySerThrTrpSerThrGlySerCysAlaThrArgCysProThrHisProSerTrp 792
Qy 1277 AGGAGCAGCGCCGTC---TGTGGACAGCAACATGCTCATCCACTTCGTGAAGAACGGGC 1333
Db 793 ThrSerCysTrpLeuSerAlaGlyAlaSerThrAlaAlaSerAlaGlyCysArgAlaSer 812
Qy 1334 CCAAGCGCTGGTGTGCTGCTCAAGTTCGTGAGCCCTGGTGTCTTCAACAAGGCCG 1393
Db 813 ProAlaHisArgGlyArgProAlaArgThrSerArgPro-----AlaArgProArgPro 830
Qy 1394 TGCTGTGGTTCGGCGCGCGTCCGCGCAAGCAGTACGCGCTGATCAAGCGCGCAGCA 1453
Db 831 AlaCys-----SerArgPheCysArgGlnProAlaThrTrp***ProArgArgAla*** 848
Qy 1454 TCC----- 1456
Db 849 AlaThrProTrpThrAlaProLeuAlaProSerArgGlyValAlaAlaValArg 868
Qy 1457 -----CCA 1459
Db 869 ProHisArgProProAlaArgProArgGlyLeuAlaProAlaHisAlaCysProProPro 888
Qy 1460 TTGAGAACTACATCGCGCACCACATGACCGCGTGGCGGAGAACT-----CGC 1507
Db 889 ThrArgProGlnSerArgAlaProArgAlaGlyAspArgGlnThrGlyValAlaArgArg 908
Qy 1508 ACCTGGCGCAAGC---AGAACTACTTCTACTACAACTGCTCAGCGCAAGTTCCTGCGCG 1564
Db 909 LeuCysAlaGlyLeuArgSerProArgAlaAlaProArgArgGlyArgProCysPro 928
Qy 1565 ACAACTGCCCCACTCTCTC----- 1585
Db 929 ThrSer-----ProGluCysArgAlaAlaGlnProGlyArgArgGlyGlyArgCysGly 946
Qy 1586 -----GCGAGCGCGCTTCGCCACCTCAAGCTGCGGTGG 1624
Db 947 ProGlyThrAlaGlyThrSerArgProProSerGlyProCysArgProArgAlaVal 966
Qy 1625 ACAACTGACCG-----TCTCCACCAACTTCTTCATGGAGGAGCTCAAGCGCGCACCT 1678
Db 967 ThrThrAlaProPheLeuGluProThrAspProAlaAlaProSerSerArgSerArg 986
Qy 1679 ACACCAAGGTGATCTGATGACACG-----TCGACTGGC 1714
Db 987 SerProArg-----SerTrpArgThrCysArgCysSerValArgSerSerTrpProGly 1004
Qy 1715 -----TCGATATGC 1723
Db 1005 GlyArgProCys***ThrArgProGlyProGlyAlaArgAlaArgValThrLeuProCys 1024

Qy 1724 CCGTGGCCAAACGAGCTGGCCGAGTGCCTGGCCCAAGCAGGTTGGCCCGCGGCGCATCGTCA 1783
Db 1025 ProAlaProTrpProArgProSerLeuGlyProAlaArgCysProLeuGlyAlaProAla 1044
Qy 1784 TCTGGCGCTCGGCTTCCTCAGCC---CGCCCTAGCCGAGCTGATCCAGAGCGGGCT 1840
Db 1045 ProProAlaProAlaProThrAlaThrArgProAlaGlyAlaTrpArgArgSerArg 1064
Qy 1841 TCGAGGTGGCTGCATCCGCGCGCCACTCAGG----- 1873
Db 1065 CysAlaCysArgSerThrGlyArgProAlaArgAlaSerArgGlnGlyProProPro 1084
Qy 1874 ---GCTACATGGACCGCGTCAACATGTACAGCTCTTCTACA 1912
Db 1085 GlySerThrAspSerThrSerAlaCysThrProThrProThr 1098

Search completed: March 14, 2006, 01:40:13
Job time : 460.5 secs

THIS PAGE BLANK (USPTO)


```
Qy 943 GGGCGCGCCAGGTGGTGTGGTGGTACTGCAACCCCGCAGTGGCGCTTCTGAGCTG 1002
Db 84 AsnProAlaSerIleAspValValAspLeuAsnProHisIleAlaLeuAsnLysLeu 103
Qy 1003 AAGAAGGTGGCCATTACAGAGCTG---GAGTTTTCAGGACGTGTGGAGCTGTTCTGGCGAG 1059
Db 104 LysLeuAlaAlaPheArgHisLeuProAlaHisGlnAspValValArgHisPheGlyArg 123
Qy 1060 GCGGTGCACCGCGCATTTAGGAGCTGTACAGAAAGAGCTGGCGCCCTCTCTGTCGCAA 1119
Db 124 AlaGlyThrArgSerAsnSerValGlyTyrAspArgPheIleAlaGluHisLeuAspAla 143
Qy 1120 ACCAGCCACAACTTCTGTCTCAAG---CCCTCTGTGTACTTC 1158
Db 144 ThrThrLysAlaTyrTrpSerLysArgThrLeuSerGlyArgArgIleSerValPhe 163
Qy 1159 CAGCAGCGCTGTACTACAGGCGGATGGGCAAGCTGTGCTGGTGTGCTGCAATGGCTG 1218
Db 164 AspArgAsnIleTyrArgThrGlyLeuLeuGlyArgPheIleGlyAlaGlyHisIleMet 183
Qy 1219 GCGGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1278
Db 184 AlaArgLeuHisGlyValLysLeuThr-----GluMetAlaLysThrArgThrLeuAsp 201
Qy 1279 GAGCAGCGCGCTGTGTGGGACAGCAATGCTCATCTTCATCTTCGTGTAAGAACGGGCCAAG 1338
Db 202 GluGlnArgGlnPhePheAspSerLysValAlaProLeuPhe-----AspLys 217
Qy 1339 CCGCTGGTG---TCGCTGTTCGTCAAGTTCGTGAGCTGTGCTCTTCAACAAGGCCGTG 1395
Db 218 ProValValArgTrpLeuThrLysArgLysSerSerLeu----- 230
Qy 1396 CTGTGTTTCGGCGCGGCTCGCGGAGCAGTAC-----GGCTGATCAAGCG 1446
Db 231 -----PheGlyLeuGlyIleProArgGlnTrpAspGluLeuAlaSerLeuSerSer 248
Qy 1447 GACGGCATCCCATTTAGAACTACATCGCGCGCACCATTGACGCGGTGGGGGAACTCG 1506
Db 249 AspGly---ThrValAlaSerValLeuLysGluArgLeuGluLysLeuAlaCysAsnPhe 267
Qy 1507 CACGTGGCAAGCAAGTACTTCTACTACAACTGCCTCACCGGCAAGTCTCTCGCGGAC 1566
Db 268 ProLeu---SerAspAsnTyrPheAlaTrpGlnAlaPheAlaArgTyrProGluPro 286
Qy 1567 AAC-----TGCCCACTACTCTGCGGAGCGGCTTCGCCCACTCAAGAGTGGC 1617
Db 287 HisGluGlyAlaLeuProAlaTyrLeuLysProGluTyrTyrGluLysIleArgAsnAsn 306
Qy 1618 GTGGTGGACAACCTGACCGCTCTCCACCACTTCTTCATGAGGAGCTCMAA----- 1668
Db 307 ThrAla---ArgValAlaValHisAlaThrTyrThrGluLeuLeuSerArgLysPro 325
Qy 1669 GCGCGCACTACCAAGGTGATTTGATGACCAACGTTGAGTCTGGTGTGATATGCCCGTG 1728
Db 326 AlaAsnGlyValAspArgTyrIleLeuLeuAspAlaGlnAspTrpMetThrAspValGln 345
Qy 1729 GCCAACGAGTGGCGAGTGTGCGCCCAAGCAGGTTCGCGCGGCGGATCGTCTCATCTGG 1788
Db 346 LeuAsnGluLeuTrpSerGlnIleSerArgThrAlaAlaSerGlyAlaArgValIlePhe 365
Qy 1789 CGCTCCGCTCC-----CTCAGCGCGCCCTACGCGGAGCTG 1824
Db 366 ArgThrAlaAlaGluLysSerValIleGluGlyArgLeuSerPro----- 380
Qy 1825 ATCCAGAGCGCGGTCTTCAGTGTGCG-----TGCAATCCGCGCGCCACT 1869
Db 381 -----AspIleArgAsnGlnTrpValTyrLeuGluGluArgSerAsn 394
Qy 1870 CAG---GGCTACATGACCGGTCAACATGTACAGCTCTTCTACATGCGCGCGG 1923
Db 395 GluLeuAsnAlaMetAspArgSerAlaIleTyrGlyGlyPheHisIleTyrGlnArg 413
```

RESULT 2

```
AG2836
conserved hypothetical protein Atu2119 [imported] - Agrobacterium tumefaciens (strain C5
C/Species: Agrobacterium tumefaciens
C/Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
C/Accession: AG2836
R/Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, E
ster, E.W.
A/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A/Reference number: AB2577; PMID:21608550; PMID:11743193
A/Accession: AG2836
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-416 <R>
A/Cross-references: UNIPROT:Q8UDK6; UNIPARC:UPI000000D1DB1; GB:AB008688; PIDN:AAL43109.1;
A/Experimental source: strain C58 (Dupont)
C/Genetics:
A/Gene: Atu2119
A/Map position: circular chromosome
Alignment Scores:
Pred. No.: 4,59e-11 Length: 416
Score: 326.50 Matches: 111
Percent Similarity: 45.1% Conservative: 69
Best Local Similarity: 27.8% Mismatches: 158
Query Match: 9.0% Indels: 61
DB: 2 Gaps: 16
US-10-620-914-44 (1-1947) x AG2836 (1-416)
```

```
Qy 823 CTGTACAGCAGTCTGGGAGGACCCGAGCGGATATGAGGTGATGGATCAACCC 882
Db 44 ValTyrProGlnIleTrpGluAspProGluIleAspMetGluAlaMetGluLeuGlyGlu 63
Qy 883 AAGGACACGCTGTGACCTGACTAGCGCGGCTGCAATGCCCTGAACTGCTGCTGCGAG 942
Db 64 GlyHisArgIleValThrIleGlySerGlyCysAsnMetLeuAlaTyrLeuSerArg 83
Qy 943 GGGCGCGCCAGGTGTGTGCTGAGTGCACAAACCCCGCAGTGGCGCTTCTGAGCTG 1002
Db 84 AsnProAlaSerIleAspValValAspLeuAsnProHisIleAlaLeuAsnLysLeu 103
Qy 1003 AAGAGTGTGCCATTACAGAGCTG---GAGTTTTCAGGACGTGTGCGAGCTGTTCTGGCGAG 1059
Db 104 LysLeuAlaAlaPheArgHisLeuProAlaHisGlnAspValValArgHisPheGlyArg 123
Qy 1060 GCGGTGCACCGCGCATTTAGGAGCTGTACAGAAAGAGCTGGCGCCCTCTCTGTCGCAA 1119
Db 124 AlaGlyThrArgSerAsnSerValGlyTyrAspArgPheIleAlaGluHisLeuAspAla 143
Qy 1120 ACCAGCCACAACTTCTGTCTCAAG---CCCTCTGTGTACTTC 1158
Db 144 ThrThrLysAlaTyrTrpSerLysArgThrLeuSerGlyArgArgIleSerValPhe 163
Qy 1159 CAGCAGCGCTGTACTACAGGCGGATGGGCAAGCTGTGCTGGTGTGCTGAGTGGCTG 1218
Db 164 AspArgAsnIleTyrArgThrGlyLeuLeuGlyArgPheIleGlyAlaGlyHisIleMet 183
Qy 1219 GCGGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1278
Db 184 AlaArgLeuHisGlyValLysLeuThr-----GluMetAlaLysThrArgThrLeuAsp 201
Qy 1279 GAGCAGCGCGCTGTGTGGGACAGCAATGCTCATCTTCATCTTCGTGTAAGAACGGGCCAAG 1338
Db 202 GluGlnArgGlnPhePheAspSerLysValAlaProLeuPhe-----AspLys 217
Qy 1339 CCGCTGGTG---TCGCTGTTCGTCAAGTTCGTGAGCTGTGCTCTTCAACAAGGCCGTG 1395
Db 218 ProValValArgTrpLeuThrLysArgLysSerSerLeu----- 230
```


Db 440 GlyAlaProGlyThrProAlaAlaProGlyPro-----GlyGlyGlyAla 454
Qy 892 CCGTGTCTTGGGTTGATCTCCA-----TCACTCCA 860
Db 455 AlaValProSerGlyAlaThrProHisProGluArgGlySerGlyProAlaAspProPro 474
Qy 859 TATCCGCTCGGGTCTCTCCACGACTCGCTGTACAGGAAGTGGGGCGGAACATGGGCG 800
Db 475 AlaAlaAlaArgLeuProProGluArgGln---GluProArgLeuProGlnAspLeuAla 493
Qy 799 GCGCTCCACGCGCTCTCTGTCAGGCGTGGCCACAGCTGGGCGAGCGCCCAATCCACA 740
Db 494 AlaAlaGlnArgCysPro-----AlaGlyProProProThr 505
Qy 739 CGTAGTAGGGGGCGCGCAGCAGCAGGACAGTATAGGGATCGAACCTGGGTCTTCTGCTCCC 680
Db 506 ArgSer---GlyAlaAlaAlaGlnArgThrHisArgArgProProGlyCysProArgSer 524
Qy 679 ACAGCGCTCCAGCTTCTGCTCCAGGTAGCGCGCGCTCGGGCCGATGTCAATGTTGT 620
Db 525 AlaArgAsnProGly----- 529
Qy 619 CGATGTCAAGATCGATCGCCAGAGAAGAACGGCGCAGCGGCA-----TCTGGC 569
Db 530 ---CysProArgThrTtpArgArgSerGlyAlaGlnArgGlyHisProProProGly 548
Qy 568 GCAGGCGAGTCTGCTACTTCCGCTCAGTGAAGTGGGCAACGCCACAGGCGCTT 509
Db 549 AlaGly-----GlnArgProSerGlyProThr 557
Qy 508 GGGACAGTACGAGCAAGCTGTGTCGATGATGCTTGTGGAACGGTGAATCATCGTAGCG 449
Db 558 GlyGlyArgProAlaAlaProGly----- 565
Qy 448 AGTAGGAAGGTGATGAGCGTCCGCGTCCAGGGGGCGCAAAATGGCAAGCTCGG 389
Db 566 -----AlaPro-----GlyThrProAla 571
Qy 388 CCTCCAGCACTGACATCTTCCAGCCCTTGGCCCTTTCCTTCTTGGCCACCTCGC 329
Db 572 AlaProGlyProGlyGlyAlaAlaVal---ProSerGlyAlaThrProHisProGlu 590
Qy 328 ACAGCGATGCGCAGGTGCGACAGTGTAGTGGACTTCCGAGGTGCGATGTAAT 269
Db 591 ArgGlySerGlyProAlaAspPro-----ProAlaAlaAlaArg 603
Qy 268 CAGCCATCATATCGACATTTCTCCCGATGTCACACCCAGGTCAACCCAGATGAGTTCC 209
Db 604 LeuPro-----ProGluArgGlnGluProArgLeuProGlnAspLeuAla 618
Qy 208 AGCCTCGG-----CCAGGGGGCAGCAAGGCGGCGCTGGGGCCGTAGAGCTCT 155
Db 619 AlaAlaGlnArgCysProAlaGly-----ProProProThrArgSerGly 633
Qy 154 CCAGCGCAGCAGCTGATCATCGC-----CCTTCTTGTCTCCGAACCAATATGGCGCA 101
Db 634 AlaAlaAlaGlnArgThrHisArgArgProProGlyCysProArgSer----- 649
Qy 100 GAACGGTCAAGTCACTCTTCATGC 77
Db 650 -----AlaArgAsnProGlyCys 655

RESULT 4

JQ0405
Hypothetical 119.5K protein (uvrA region) - Micrococcus luteus
N:Alternate names: ORF 1 protein
C:Species: Micrococcus luteus, Micrococcus lysodeikticus
C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 04-Feb-2000
C:Accession: JQ0405
R:Shiota, S.; Nakayama, H.
Mol. Gen. Genet. 217, 332-340, 1989
A:Title: Micrococcus luteus homolog of the Escherichia coli uvrA gene: identification of
A:Reference number: S04781; MUID:89364717; PMID:2549377

A:Accession: JQ0405
A:Molecule type: DNA
A:Residues: 1-1106 <SHI>
A:Cross-references: UNIPARC:UPI0000177394; EMBL:X15867
A:Note: this reading frame extends between two stop codons and does not begin with a start
A:Note: the gene encoding this protein overlaps uvrA gene
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
Alignment Scores:
Pred. No.: 2,51e-10 Length: 1106
Score: 315.00 Matches: 208
Percent Similarity: 33.4% Conservative: 35
Best Local Similarity: 28.6% Mismatches: 231
Query Match: 8.7% Indels: 254
DB: 2 Gaps: 45

US-10-620-914-44 (1-1947) x JQ0405 (1-1106)

Qy 273 CATCGACCTGGCGAAGTTCAAGTCCATCTACGTGGTGGCTGACCTGTGCCACTCGCTGTGCGA 332
Db 300 HisArgLeuGlyGlyAspArgProGluAlaGlyArgProArgArgGlyArgValArg 319
Qy 333 GGT-----GCCCAAGAAAGGC-----GAAGCCCAAGGCTGGAGAAATGTCCA 377
Db 320 GlyArgGlyAlaGlyArgArgGluGluHisArgGlyValArgAlaGly 339
Qy 378 GGTGTGGAGCGCCGACGCTTGCCTTGCCTTGC----- 407
Db 340 ArgArgGlyGlnProProValProLeuValLeuArgGluAlaLeuLeuProGlnArgAla 359
Qy 408 ---GCCCTTGGAGCCAGCGCTCATCATCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 464
Db 360 ArgAlaAspArgGly---ArgAspArgAlaAlaLeuValLeuLeuGlnProValArg 378
Qy 465 ACCGTTCACAAAGCTCATCGA-----CCAGGCTTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 515
Db 379 ArgValProArgValHisArgHisArgLeuProAlaGlyGlyProGlyProArgArg 398
Qy 516 -----CCTGTGGCGCTTGGCGCTTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 554
Db 399 ArgGlnArgAlaValProAlaArgGlyArgArgAlaValAlaAlaArgGlnVal 418
Qy 554 ----- 554
Db 419 HisLeuGlyLeuLeuAlaAlaArgAlaArgAlaGlyGlnGlyAspGlyLeuLeuPro 438
Qy 555 ---CGACTCGCTGGCGCAGATGCCCTGGTGGCGCTTCTTCTGTGGCGATCGATCTT 611
Db 439 GlyHisProValGluGlyProAspGlyGlyAla----- 450
Qy 612 CGACATCGACACATTGACAT-----CGG 635
Db 451 ArgArgArgProAlaArgGlnGlyLeuGlnGlyGlyGlyAspValProGlnProValArg 470
Qy 636 CCGCGAGCGCG-----CGCTCTGTGGAGCAAGCT----- 668
Db 471 ProArgAlaProLeuHisHisGlyLeuArgGlyArgHisProLeuArgAspAlaGlnAla 490
Qy 669 ---GGAGCGGTGTGGGAGCAGAA-----CACCAGGCTTGCAT 704
Db 491 ArgGlyAspArgValGlyArgArgProArgProArgAlaLeuArgValHisAlaGly---Asp 509
Qy 705 CCTCTACGTGGTGGTGGCGCGCCCTACTACTGTGTGATTTGGCTGTGGCTGTGGCGCGT 764
Db 510 ProValProGlyValProArg-----GlyProProGlnProHis 522
Qy 765 TGGCCACCGCTCGCACGAGAGCG-----CGTGAGCGCGCGCGCTTGTTCGCCGCC 815
Db 523 GlyProGluArgAlaArgGlyArgProValHisArgGly-----ArgHisProProAla 540
Qy 816 CACCTTCTGTACAGCAGCTGTGGGAGGAGCCCGAGCGGATATGGAGGTGTGGAGAT 875
Db 541 HisAlaArgGlyHisGlyValLeuLeuGly-----AlaAlaAlaAspGlyProGlyAla 558

Db 117 ProLeuLeuAlaMetAlaThrThrCysCysGlyValArgProTrpAlaTrpHisSerThr 136
Qy 473 ACACAGTCATCGACCGCTTGTCTGCTACCTGCTCCCAAGACGGCTGCTGGCGGTGCGCG 532
Db 137 GlyAsnSerArgCysThrValSerSerAlaCys-----LeuPro 149
Qy 533 ACTTCTACGTGAGGGGCAAGTACGACCTGCGCCCTGCGCCAGATGCTGCTGCGCGCGTT 592
Db 150 ProSerTrpLeuAlaMetAlaThrThrCysSerGlyValArg---ProTrpAlaTrpHis 168
Qy 593 TCTTCTGGCATCGATCTTCGACATCGACACATGACATCGCCCGCGAGCGCGCGCT 652
Db 169 SerThrGly-----AsnSerArgCysThrValSerSerAlaGlyLeuProPro 185
Qy 653 ACCTGGAGCAGACCTGGAGCGGTGTGGAGCAGA-----ACACCC 694
Db 186 LeuLeuAlaMetAlaThrThrCysCysGlyValArgProTrpAlaTrpHisSerThrGly 205
Qy 695 AGGTTTCGATCCCTACGTGCGGTGCGCGCTGCGCGCCCTACTACGTGTGGATTGGCGGCC 754
Db 206 AsnSerArgCysThrValSerSerAlaCysLeuProProSerTrpLeu----- 221
Qy 755 TGCCACAGCTTGGCCAGCCCTGCACGAGGAGC-----GGTGGAGCGCGCGCCCA 805
Db 222 -----AlaMetAlaThrThrCysSerGlyValArgProTrpAlaTrpHisSerThrGly 239
Qy 806 TGTTCGCGC-----CCACCTTCC-----TGACA 829
Db 240 AsnSerArgCysThrValSerSerAlaCysLeuProProSerTrpLeuAlaMetAlaThr 259
Qy 830 CGCAGTCGTGGAGGACCGCGAGCGGATGAGGTGAGTGGAGATCAACCCCAAGGACA 889
Db 260 ThrCysSerGlyValArgPro-----TrpAla---TrpHisSerThrGlyAsnSer 275
Qy 890 CGTGTGCTGACCTGACCTAGCGCGCTGCAATGCCCTGAACCTCTGCG----- 937
Db 276 ArgCys-----ThrValSerSerAlaCysLeuPro---ProSerTrpLeuAlaMetAla 292
Qy 938 -----TGCAGGGGCGCGCAGGTGTGTGCTGGTGGAGTGCACCCCGCGAGTGGCGCG 991
Db 293 ThrThrCysSerGlyValArgProTrpAla---TrpHisSerThrGlyAsnSerArgCys 311
Qy 992 TTTCTGGAGC-----TrpAla---TrpHisSerThrGlyAsnSer 1000
Db 312 ThrValSerSerAlaCysLeuProProLeuLeuLeuAlaMetAlaThrThrCysSerGly 331
Qy 1001 TGAAGAAGGTGGCCATTGACGAGCTGGAGTTTGAGGACGTGTGGCAGCTGTTCGGCGAGG 1060
Db 332 ValArgProTrpAlaTrpHisSerThrGlyAsnSerAlaCysThrValSerSer----- 349
Qy 1061 GCGTGCACCGCGCATTTG-----AGGAGCTGTACGAGA---AGNAGCTGG 1102
Db 350 AlaCysLeuProProLeuLeuLeuAlaMetAlaThrThrCysSerGlyValArgProTrp 369
Qy 1103 CGCCCTTCTGTCGCAACACGACCACTTCTGCTCCAGCGCTCTGCTGCTTCCAGC 1162
Db 370 -----AlaTrpHisSerThrGlyAsnSerArgCysThrValSer 382
Qy 1163 ACGGCTGTACTACGAGGCGCATGGGCAAGCTGTGCTGGGTGCTGCTGCTGCTGGCGCG 1222
Db 383 SerAlaCysLeuProProLeuLeuAlaMetAlaThrThrCysCysGlyValArgPro 402
Qy 1223 TGGTGTGG----- 1231
Db 403 TrpAlaTrpHisSerThrGlyAsnSerArgCysThrValSerSerAlaCysLeuProPro 422
Qy 1232 GACTGG---GCAAGACCGTCAAGCGCTCGCCCAAGCGGCCCA---CAATGGAGGAGCAGC 1285
Db 423 SerTrpLeuAlaMetAlaThrThrCysSerGlyValArgProTrpAlaTrpHisSerThr 442
Qy 1286 GCCGCTGTGGGACGACCAATGCTCATCTTCATCCACTTGTGAAGACGCGGCCCAAGCGCTGG 1345

Db 443 -----GlyAsnSerArgCysThrValSerSerAlaCysLeuProProSer---Trp 458
Qy 1346 TGTGGCTGTTCTGCAAGTTCTGAGCTGCTGCTCTTCAACAGGCGC----- 1393
Db 459 LeuAlaMetAlaThrThr-----CysSerGlyValArgProTrpAlaTrpHis 474
Qy 1394 -----TGCTGTGTTCTGCGGCGCGCTGCGCGCAAGCAGTACGCGCTGA 1438
Db 475 SerThrGlyAsnSerAlaCysThrValSerSerAlaCysLeuProProLeuLeuAla 494
Qy 1439 TCAAGGGGAGCGCATCCCATTTGAGAACTATACGCGCGCACCATGAGCGGTGGCGG 1498
Db 495 MetAlaThrThrCysSerGlyValArg-----ProTrp---AlaTrpHis 508
Qy 1499 AGAACTCGCAGCGGCGCAAGCAGAACTACTTCTACTACAACTGCCTCACCG----- 1549
Db 509 SerThrGlyAsnSerAlaCys---ThrValSerSerAlaCysLeuProProLeuLeu 527
Qy 1550 -----GCAAGTTCTCTCGCGCAACTGCGCCA----- 1576
Db 528 AlaMetAlaThrThrCysSerGlyValArgProTrpAlaTrpHisSerThrGlyAsnSer 547
Qy 1577 ---CCTACCTGCGCGAGGCGCTTCCGCCACCTCAAGAGTGGCTGGTGGACAACTGA 1633
Db 548 AlaCysThrValSerSerAlaCysLeuProProLeuLeuAlaMetAlaThrThrCys 567
Qy 1634 CCGTCTCCCAACTTCTTCATGAGGAGCTCAAGCGCGCA----- 1675
Db 568 SerGlyValArgProTrpAlaTrpHisSerThrGlyAsnSerArgCysThrValSerSer 587
Qy 1676 -----CCTACACACAGGTGATTCTGTATGGACCACTGGTGGTGGTATGCGCG 1726
Db 588 AlaCysLeuProProLeuLeuAlaMetAlaThrThrCysSerGly---ValArgPro 606
Qy 1727 TGCCCAACGAGCTGG----- 1741
Db 607 Trp-----AlaTrpHisSerThrGlyAsnSerArgCysThrValSerSerAlaCysLeu 624
Qy 1742 -----CCGAGTGTCTGGCCA----- 1756
Db 625 ProProLeuLeuAlaMetAlaThrThrCysCysGlyValArgProTrpAlaTrpHis 644
Qy 1757 AGC-----AGGTGGCGCGCGCATCTGCTATCTGGCGCTCCGCTCC----- 1801
Db 645 SerThrGlyAsnSerArgCysThrValSerSerAlaCysLeuProProLeuLeuAla 664
Qy 1802 -----TCAGCGCGCT----- 1813
Db 665 MetAlaThrThrCysCysGlyValArgProTrpAlaTrpHisSerThrGlyAsnSerArg 684
Qy 1814 -----ACGCGAGCTGTATCCAGAGGCGGCTTCGACGTGCGCT 1852
Db 685 CysThrValSerSerAlaCysLeuProProSerTrpLeuAlaMetAlaThrThrCysSer 704
Qy 1853 -----GCATCCGCGCGCCACTCAGGCTACATGAGCGGTCAACA 1894
Db 705 GlyValArgPheArgGlyAlaProValSerAlaLeuArgHisAlaCysThrAlaSerThr 724
Qy 1895 TGTACAGCTCT-----TCTACATGGCCC 1918
Db 725 AlaValThrProProArgPheAlaIleLeuArgSerTrpPro 738

RESULT 6
T43481
probable mucin DKFP434C196.1 - human (fragment)
N/Alternate names: protein DKFP434B0635.1
C/Species: Homo sapiens (man)
C/Date: 21-Jan-2000 #sequence revision 21-Jan-2000 #text_change 09-Jul-2004
C/Accession: T43481; T17264
R/Koehler, K.; Beyer, A.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, December 1999
A/Reference number: 222514
A/Accession: T43481

A:Molecule type: mRNA
 A:Residues: 1-580 <AAA>
 A:Cross-references: UNIPROT:Q9UF83; UNIPARC:UPI000006D400; EMBL:AL133561; NID:g6599133;
 R:Experimental source: adult testis; clone DKFZp434C196
 R:Poustka, A.; Wellenreuther, R.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
 submitted to the Protein Sequence Database, October 1999
 A:Reference number: Z11540
 A:Accession: T34549
 A:Molecule type: mRNA
 A:Residues: 262-580 <POU1>
 A:Cross-references: UNIPARC:UPI0000070F36; EMBL:AL122069; NID:g6102864; PIDN:CAB59245.2;
 R:Experimental source: adult testis; clone DKFZp434B0635
 R:Poustka, A.; Klein, M.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
 submitted to the Protein Sequence Database, September 1999
 A:Reference number: Z18723
 A:Accession: T17264
 A:Molecule type: mRNA
 A:Residues: 262-580 <POU2>
 A:Cross-references: UNIPARC:UPI0000070F36; EMBL:AL117481; NID:g5911958; PIDN:CAB59594.1;
 R:Experimental source: adult testis; clone DKFZp434B061
 C:Genetics:
 A:Note: DKFZp434C196.1; DKFZp434B0635.1; DKFZp434B061.1

Alignment Scores:

Pred. No.: 1.38e-08 Length: 580
 Score: 283.50 Matches: 190
 Percent Similarity: 36.1% Conservatives: 52
 Best Local Similarity: 28.3% Mismatches: 260
 Query Match: 7.8% Indels: 169
 DB: 2 Gaps: 31

US-10-620-914-44 (1-1947) x T43481 (1-580)

QY	8	CGGGTGTGACGGCGCGCTGCGAGCTACACCAAGAAAGTCTTCCTCGGAGAGCTCA	67
DB	19	ArgProSerThrAlaSerLeuThrArgThrProSerArgAlaSerProThrArgMetPro	38
QY	68	AGCTCAGCAGCAAGAGTACCTGACCGTTCCTGCGCCATATGTGGTTCGGCAGCAAGA	127
DB	39	SerArgAlaSerLeuLysMetThr---ProPheArgAlaSerLeuThrLysMetGluSer	57
QY	128	AGGGGATGATCAGCTGCTCGCTGGAGAGCTTCT	163
DB	58	ThrAlaLeuLeuArgThrLeuPro---ArgAlaSerLeuMetArgThrProThrArgAla	76
QY	164	-----ACGGGCCCGACGCGCTGCTTTG---CTGCCGCTGGCGAGGCT	208
DB	77	SerLeuMetArgThrProProArgAlaSerProThrArgLysProProArgAlaSerPro	96
QY	209	CGAACCTCATCTGGGTTGACCTGGGTGGTGGCCTGGGAGAAATGTCGATATGAGCTG	268
DB	97	ArgThrProSer-----	100
QY	269	ATTACATCGACCTGGGGAAGTCAAGTCCATCTACGTGTGTCGACCTGTGCCACTCGCTGT	328
DB	101	-----ArgAlaSerProThr-----	105
QY	329	GCAGAGTGGCAAGAAGCGCAAGCGCAAGCGCTGGAAGAATGCCAGTCTGGAGG	388
DB	106	-----ArgArgLeuProProArgAlaSerProMetGlySerProHisArg	119
QY	389	CCGACGCTTCCCAATTTGCGCCCTGAGGGCACCGCGCTCATCACCTTCTCTCTACT	448
DB	120	AlaSer-----ProMetArgThrProProArgAlaSerProThrGlyThr	134
QY	449	CGCTCAGGATGATCCACCGTTCCACACGTCATTCGACACGAGGCTGTGCTGCTACTGCTCC	508
DB	135	ProSerThrAlaSerProThrGlyThrProSerSerAlaSerProThrGlyThrProPro	154
QY	509	AAGACGCGCTGGTGGGCTTCCGACTTCTACGTGAGCGCAAGTACGACCTGCCCTGCG	568
DB	155	Arg-----AlaSerProThrGlyThr-----	161
QY	569	GCCAGATGCCCTGGTCCGCGCTTCTCTGCGGATCGATCTTCGACATCGACAACATTG	628
DB	162	-----ProProArgAlaTrpAlaThr-----ArgSerProSerThrAlaSerLeu	176
QY	629	ACATCGCGCCCGAGCGCGCTTACTTGAGCAGAAAGCTGGAGCGGTGGGAGCAGA	688
DB	177	ThrArgThrProSerArgAlaSerLeuThrArgTrpProProArgAlaSerProThrArg	196
QY	689	ACACCCAGGTTCCGATCCCTACGTCCGCTGGCTGGCGCCCTCTACTACGTGTGGATTG	748
DB	197	ThrProProArgGluSerProArgMetSerHisArgAlaSerProThr-----	212
QY	749	GCGCGCTGCCAGCGTTGGCCAGCGCTGCACGAGGAGCGGTGGAGCGCGCCCATGT	808
DB	213	---ArgThrProProArgAlaSerPro---ThrArgArgProProArgAlaSerProThr	230
QY	809	TCCCGCCCACT-----TCCTGTACACGAGTCGTGGGAGGACCCGAGCCGATATGG	862
DB	231	ArgThrProProArgGluSerLeuArgThrSerHisArgAlaSerProThrArgMet---	249
QY	863	AGGTGATGAGATCAACCCCAAGGACACGCTGCTGACCTGACTAGCGCGGTGCAATG	922
DB	250	-----ProProArg-----	254
QY	923	CCCTGAACCTGTGTGCGAGGGGCGCGCAGCGTGTGTGCGTGAGCTGTACGAGAAGCTGG	982
DB	255	Pro-----ThrArgArgProProArgAlaSerProThrGlySerProProArg	270
QY	983	AGTCGGCGCTTCTGGAGCTGACAGAGGTGGCCATTGACGAGCTGGAGTTGAGGACGTGT	1042
DB	271	AlaSerProMetThrProProArgAlaSerPro-----ArgThrPro	284
QY	1043	GGCAGCTGTTTCGCGAGGCGCTGCACCCCGCATTCGAGGAGCTGTACGAGAAGCTGG	1102
DB	285	ProArgAlaSerPro---ThrThrThrProSerArgAlaSerLeuThrArg-----	300
QY	1103	CGCCCTTCTGTGCGAAACAGCCCAACT---TCTGGTCCAAGCGCTCTGTACTTCC	1159
DB	301	ThrProSerTrpAlaSerProThrThrProSerArgAlaSerLeuMetLysMetGlu	320
QY	1160	AGCAGCGCTGTACTACGAGGCGGCTGGCAAGCTGTCTGGTGTCTGCTGCTGCTGCTG	1219
DB	321	SerThrValSerIleThrArgThrPro-----	329
QY	1220	CGTGTGTGCTGGAGCTGGCGAAGCGCTCAAGCGCTCGCCACACGCGCCCAATGAGG	1279
DB	330	ProArgAlaSerProThrGlyThrProSerArgAlaSerProThrGlyThrProSerArg	349
QY	1280	AGCAGCGCTGTGTGGG-----ACAGCAACATGCTCATCCACTCG	1321
DB	350	AlaSerLeuThrGlySerProSerArgAlaSerLeuThrGlyThrProSerArgAlaSer	369
QY	1322	TGAGAACGGGCGCCAGCGCTGGTGGCTGTTCGTCAGTTCTGAGCGCTGGTCTCT	1381
DB	370	LeuIleGlyThrProSerArgAlaSerLeuIleGlyThrProSerArgAlaSerLeuThr	389
QY	1382	TCAACAAGCGCGTGTGTGTGCTGGCGCGGCTGCGCGGAGCTAGTCGCTGCTCA	1441
DB	390	GlyThrProPro-----ArgAlaSerLeuThrGly---Thr	400
QY	1442	AGCGGACGGCATTCCTATGTAGAACTATATCGCGCGCACTATGAGCGCGGTGGGAGA	1501
DB	401	SerSerThrAlaSerLeuThrArgThrProSerArgAlaSerLeuThr-----	416
QY	1502	ACTCGCAGCTGGCAGCAGAACTACTTCTACTACACTGCCTCAGCGGCAAGTCTCTGC	1561
DB	417	-----ArgThrGlnSerSerSerLeuThrArgThrProSerMetAlaSer	432
QY	1562	GCAGCAACTGCCCACTACCTGCGGAGCGCGCTTCGCCACCTCAAGAGGTGGGTGG	1621
DB	433	LeuThrArgThrProProArgAlaSerLeuThrArgThrProPro---ArgAlaSerLeu	451
QY	1622	TGACAACTGACCG-----TCTCCACA-----	1645


```
QY 1523 ACTACTTCTACTCAACTGCTCACC- - - - -GCAAGTTCTCGCGCACTGCGCCCA 1576
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 390 AlaThrAlaAlaThrAlaAlaThrProAlaAArgAlaAlaThrProAlaThr 409
QY 1577 CTACTCTCGCGAGCGCGCTCGCCACCTCAAGAGTGGCGGTGGACAACTGACCG 1636
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 410 ProAlaThrProAlaThrProAlaThrProAlaThrAlaAlaThrAlaAla 429
QY 1637 TCTCCACCAACTTCTTCATGG- - - - -AGGAGCTCAAGCGCGCACCTACACCA 1684
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 430 ThrAlaAlaThrAlaAlaThrAlaAlaThrAlaAlaThrAlaAlaThr 448
QY 1685 AGGTGATCTGATGGACACGCTGGAGTGGCTGGATATGCCCGTGGCCACAGCTGGCCG 1744
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 448 - - - - - 448
QY 1745 AGTGCCTGGCCAGCAGGTGGCGCGGCGCATCGTCATCTGGCGCTCGCCCTCCCTCA 1804
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 449 - - - - -ProAlaAArgAlaAlaAArgAlaAlaThrProAlaThrGlyAlaThrProAla 465
QY 1805 GCCCGCCCTACGCGGAGCTGATCCAGAAAGCGCGCTTCGACGTGCGTCATCGCGCGCG 1864
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 466 ThrAlaProThrAlaGly- - - - -ThrAlaAlaThrAlaAlaThrAlaAlaThrAla 484
QY 1865 CCACTCAGGCTACATGGACCGCGTCAACATGTACAGCTCTCTTCTACA 1912
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 485 ProAlaAArgAlaSer- - - - -ThrProAlaThrAlaAlaThrProValThr 499

RESULT 8
S27923
gene LF3 protein - human herpesvirus 4
C:Species: human herpesvirus 4, Epstein-Barr virus
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 09-Jul-2004
C:Accession: S27923
R:Parker, B.D.; Bankier, A.; Satchwell, S.; Barrell, B.; Farrell, P.J.
submitted to the EMBL Data Library, August 1990
A:Description: Sequence and transcription of Raji Epstein-Barr virus DNA spanning the B9
A:Reference number: S27923
A:Accession: S27923
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-924 <PAR>
A:Cross-references: UNIPROT:Q99307; UNIPARC:UPI00000615C0; EMBL:M35547; NID:g330420; PID
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;

Alignment Scores:
Pred. No.: 1,31e-07 Length: 924
Score: 267.00 Matches: 196
Percent Similarity: 30.8% Conservative: 40
Best Local Similarity: 25.6% Mismatches: 296
Query Match: 7.4% Indels: 234
DB: 2 Gaps: 41

US-10-620-914-44 (1-1947) x S27923 (1-924)
QY 1888 CGCGGTCCATGT- - - - -AGCCCTCAGTGGCGCGCGGA 1856
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 6 ArgGlyProCysLeuAlaProGlySerGlyLeuGlyAlaHisProHisProArgArgSer 25
QY 1855 TGACGCGCAGCTGCAAGCCGCGCTTCTGGA- - - - -TCAGTCCGCGTAGGGCGGCTGA 1802
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 26 GlyAlaAlaAspProAlaAspProValGlyHisProAlaAlaProArgAlaProGly- - - 44
QY 1801 GGGAGCGGAGCGCCAGATGACATGCCCGCGCGCACTCTGCGCCAGCATCGG 1742
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 45 - - - - -ProGluProArgThrArgLeuGlnProAlaThrProArgArgSerGlyAlaAla 62
QY 1741 CCAGCTCGTGGCCAGCGCATATCCAGCCAGTCCACGTGGTCCATCAGAAATCACCTTGG 1682
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 63 AspProAlaAspProValGlyHisProAlaAlaProArgAlaProGlyProGluProArg 82
QY 1681 TGTAGTGGCGCGCTTTGAGCTCTCCATGAAGAAGTTGTGGAGACGGTCAAGTTGT- - - 1625
```

```
Db 83 ThrArgLeuGlnProAlaThrPro- - - - -ArgArgSerGlyAlaAla 96
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
QY 1624 - - - - -CCACCACGCCACTCTTGAGGGTGGCGAAGCGC- - - - -CCTCGC 1586
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 97 AspProAlaAspProValGlyHisProAlaAlaProArgAlaProGlyProGluProArg 116
QY 1585 GCAGGTAGTGGGGCAGTTGTTCGCGCAGAACTTGCCTGGAGCAGTGTGTAGTAGAAGT 1526
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 117 Thr- - - - -ArgLeuGlnProAlaThrProArg 125
QY 1525 AGTTCTGCTTGGCAGCTGGAGTTCTCGCCACAGCGCTGCATGTGGCGCGCATGTAGT 1466
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 126 ArgSerGlyAlaAlaAspProAlaAspProValGlyHisProAlaAlaProArgAlaPro 145
QY 1465 TCTCAATGGGATGCCGCTCGCCTTGCATCAGCGCTACTGCTTGGCCGCGCACCGCCGCGC 1406
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 146 - - - - -GlyProGluProArgThrArgLeuGlnProAlaThrPro- - - - -ArgArg 160
QY 1405 CGAACACACAGCAGCGCT- - - - -TGTGAAGAGACACACAGC 1370
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 161 SerGlyAlaAlaAspProAlaAspProValGlyHisProAlaAlaProArgAlaProGly 180
QY 1369 TCACGAACTTGACCAACACCCACAGCGCTTGGCGCGCTTCTTCACGAAGTGGATGA 1310
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 181 ProGluProArgThrArgLeuGlnProAlaThrProArgArgSerGlyAlaAlaAsp- - - 199
QY 1309 GCATGTTGCTGCCACAGCAGCGCTGCTCTCCATTTGGCGCGCTTGGCGAGCGCT 1250
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 200 - - - - -ProAlaAspProValGlyHisProAlaAlaProArg- - - - - 211
QY 1249 TGACGGTCTTGGCCAGTCCACAGCACACCG- - - - -CCAGGCATCTGACGACCC 1202
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 212 - - - - -AlaProGlyProGluProArgThrArgLeuGlnProAlaThrProArgArg 228
QY 1201 AGCAGAGCTTGC- - - - -CCATCGCGCTTGTGTAGTACAGGC 1166
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 229 SerGlyAlaAlaAspProAlaAspProValGlyHisProAlaAlaProArgAlaProGly 248
QY 1165 CGTGTCTGAAGTACCAGA- - - - -GGCGCTTGGACCAAGTTGTGGCTTGTTCGCACA 1112
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 249 ProGluProArgThrArgLeuGlnProAlaThrProArgArgSerGly- - - - -AlaAlaAsp 267
QY 1111 GGAAGGGCCCGACTTCTTCGTACAGCTCTCAATGCGCGGT- - - - - 1067
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 268 ProAlaAspProValGlyHisProAlaAlaProArgAlaProGlyProGluProArgThr 287
QY 1066 - - - - -GCAGCGCTCGCCGACAGCTGCCACAGCTCTCAAACTCCAGCTGCT 1019
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 288 ArgLeuGlnProAlaThrProArgArgSerGlyAlaAlaAspProAlaAspProValGly 307
QY 1018 GAATGGCCACTTCTTCAGCTCCAGAAAGCGCGACTCGCGGGTTCAGTCCACCGCACA 959
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 308 His- - - - -ProAlaAlaProArgAlaProGlyProGluProArgThr 321
QY 958 - - - - -CCACTGCG 950
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 322 ArgLeuGlnProAlaThrProArgArgSerGlyAlaAlaAspProAlaAspProValGly 341
QY 949 CGGCCCTCGCACCGAGCTTCAGGGCATTCGACGCC- - - - - 911
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 342 HisProAlaAlaProArgAlaProGlyProGluProArgThrArgLeuGlnProAlaThr 361
QY 910 - - - - -CGTACTCAGGGTCACA- - - - -CGTGT- - - - -CCTTGGGTGTCTCCATCACCTCCA 860
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 362 ProArgArgSerGlyAlaAlaAspProAlaAspProValGlyHisProAlaAlaProArg 381
QY 859 TATCCGGCTCGGGTCTCCACG- - - - -ACTCGCTGTACAGAAAGTGGCGGGAACA 806
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 382 AlaProGlyProGluProArgThrArgLeuGlnProAlaThrProArgArgSerGlyAla 401
QY 805 - - - - -TGGCGCGCTCCAGCGCTCTCGTGAGCGCGT 770
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
```

Db 402 AlaAspProAlaAspProValGlyHisProAlaAlaProArgAlaPro----- 417
 Qy 769 GGCAACGCTGGGAGCGGCCAATCCACAGTAGTAGGGGGCGCGACCGACGACGT 710
 Db 418 -----GlyProGluProArgThrArgLeuGlnProAlaThrProArg 431
 Qy 709 AGGGATCGAACCTGGGTGTTGCTCCACACGCGCTCCAGCTTCTGCTCCAGGTAGG 650
 Db 432 ArgSerGlyAlaAlaAspProAlaAspProValGlyHisProAla---AlaProArgAla 450
 Qy 649 CGGGCGCTCGGGCCGATGTCATGTGTGCGATGTCGAAGATCGATCGCCAGAAGAAC 590
 Db 451 ProGlyProGluProArg-----ThrArgLeuGlnProAlaThrProArgArgSer 467
 Qy 589 GGGCGACACAGGCATCTGGCGAGGGGAGGCTGCTACTTGGCGCTCACGTAGATCGG 530
 Db 468 GlyAlaAlaAspProAlaAspPro---ValGlyHisProAlaAlaProArgAlaProGly 486
 Qy 529 CAACGCCACCGAGCGCTCTGGGACAGGTACGAGCAAGCTGCTGATGACGTTGGGA 470
 Db 487 ProGluProArgThrArgLeuGlnProAlaThrProArg-----ArgSerGly 502
 Qy 469 ACGTGTGAATCATCTGAGCGAGTAGGAGAGGTGATGAGCGTGGCGTCCCTCAGGGG 410
 Db 503 AlaAlaAspProAla-----AspProValGly 511
 Qy 409 GCGCAATTTGGCAAGCTCGGCT---CGAGGACCT----- 377
 Db 512 HisProAlaAlaProArgAlaProGlyProGluProArgThrArgLeuGlnProAlaThr 531
 Qy 376 -----GlyProGlu-----GGACATCTTCCAGCCCT--- 359
 Db 532 ProArgArgSerGlyAlaAlaAspProAlaAspProValGlyHisProAlaAlaProArg 551
 Qy 358 TGGCTTTCGGCTTCTTCTGGCCACCTCGCACGCGAGTGGCACAGGTTCGACCGCTAGA 299
 Db 552 AlaProGlyProGlu-----ProArgThrArgLeuGlnProAlaThrProArgArg 568
 Qy 298 TGGACTTGAATCTGCGCAGGTGCGATGTAATCAGCCATCATCGACATCTCCCGAGTGC 239
 Db 569 SerGlyAlaAlaAspPro-----AlaAspProVal 578
 Qy 238 CACACCCAGGTCAACCCAGATGAGTTTCGAGCGCTCGGCGAGCGGGCAGCAAGGACAG 179
 Db 579 GlyHisProAlaAlaProArg-----AlaProGlyProGluProArgThr 593
 Qy 178 CGG----- 176
 Db 594 ArgLeuGlnProAlaThrProArgArgSerGlyAlaAlaAspProAlaAspProValGly 613
 Qy 175 ---CCTGGGGCCGTAGAAGCTCTCCAGGCGAGCGGTGATCAT---CGCCCTTCTTGC 122
 Db 614 HisProAlaAlaProArgAlaProGlyProGluProArgThrArgLeuGlnProAlaThr 633
 Qy 121 TGCGGAACCATATGCGCGAGAACGCTCA-----GTCATCTTCTATGCTGCTGA 71
 Db 634 ProArgArgSerGlyAlaAlaAspProAlaAspProValGlyHisProAla----- 650
 Qy 70 GCTTGAAGCTTCTCCAGGG-----AGAAAGTTCTTCTTGGTGTAGTCTGCAG 26
 Db 651 AlaProArgAlaProGlyProGluProArgThrArgLeuGlnProAlaThrProArgArg 670
 Qy 25 GCGGGCGTCCAGACCG 8
 Db 671 SerGlyAlaAlaAspPro 676

RESULT 9

F75518

hypoetical protein - Deinococcus radiodurans (strain R1)

C;Species: Deinococcus radiodurans

C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004

C;Accession: F75518

R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;

M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Mal
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
 Science 286, 1571-1577, 1999
 A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
 A;Reference number: A75250; MUID:20036896; PMID:10567266
 A;Accession: F75518
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-839 <WHI>
 A;Cross-references: UNIPROT:Q9RX57; UNIPARC:UPI00000C1761; GB:AE001904; GB:AE000513; NID
 A;Experimental source: strain R1
 C;Genetics:
 A;Gene: DR0458
 A;Map position: 1
 C;Superfamily: herpesvirus immediate-early protein IE110; RING finger homology

Alignment Scores:

Pred. No.: 1.68e-07 Length: 839
 Score: 265.00 Matches: 157
 Percent Similarity: 31.8% Conservative: 51
 Best Local Similarity: 24.0% Mismatches: 242
 Query Match: 7.3% Indels: 204
 DB: 2 Gaps: 27

US-10-620-914-44 (1-1947) x F75518 (1-839)

Qy 1771 CCGCGCAACCTCTTGGCCA-----GGCACTCGGCAGCTCGTTGGCCACGGCA 1721
 Db 143 ProAlaGluProLeuLysProValGlnAspThrProProValThrProLysPro 162
 Qy 1720 TATCCAGCCAGTCCAGTGGTCCATCAGATCATCATGTGTGTAGTGTGGCGCTTTGAGCT 1661
 Db 163 ValThrProGluProValThrProLysProAlaPro-----Thr 175
 Qy 1660 CTTCCATGAAGATTGTGTGGACAGCGTCAAGTTGTCCACACGCCACTCTTGAGGTGG 1601
 Db 176 ProGluValLeuGlnProProValAlaGlnThrProPro-----Val 190
 Qy 1600 CGAAGCGCGCTCGCGCAGGTAGTGGGCGAGTTGTTCGCGCAGGAACCTTCCGCTGAGGC 1541
 Db 191 AlaLysProPro----- 194
 Qy 1540 AGTTGTAGTAGAAGTAGTTCTGCTTGGCAGTCGCGAGTTCTCGCCACGCCCTCCATGG 1481
 Db 195 -----ValProAlaProThrSerGlnThrProThrProValGlnProAla 210
 Qy 1480 TGGCGCGAGTAGTTCTCAATGGGATGCGCTCGCTTGTATCA-----GCGGTACT 1427
 Db 211 ProThrArgThrProProGlnAlaAlaArgProThrProAsnAlaProAlaGlnThr 230
 Qy 1426 GCTTGGCCGCGCAGCGCCGCCGACACACAGCAGCGCTTGTGAAGACAGCAGGCTCA 1367
 Db 231 ProAlaProAlaThrGlnAlaProAlaAlaGlnThrProThrAlaGlnAlaProAlaThr 250
 Qy 1366 CGAATCTGACGAACAGCACACACAGCGCTTGGGCCCGTTCTTCACGAGTGGATGAGCA 1307
 Db 251 GlnThrProAlaThrProAlaProAlaAlaGlnArg----- 262
 Qy 1306 TGTGTGTCCTCCACAGCGCGTGTCTCTCCATGTGTGGCGCGTGTGGCGAGCGCTTGA 1247
 Db 263 -----ProAlaGlyGlyAla----- 267
 Qy 1246 CGGTCTTGGCCAGTCCCGCAGCAGCCAGCGCCAGGCACTGCA----- 1208
 Db 268 -----ProSerProAlaProAlaProAlaGlnAlaAsnAlaProAlaGlySerVal 284
 Qy 1207 -----GCACCCAGCACAGCTTGGCCCTTGGCCCTGTGTGTACAGGCGCT 1163
 Db 285 ValProGluAlaThrValProGluSerSerThrProAlaAlaProSerAlaGlnThrPro 304
 Qy 1162 GCTGGAAGTCCAGAGCGCTTGGACCAAGATGTGTGCTGTGTGGACAGGAGGGCG 1103
 Db 305 ProThrProThrArg-----GluThrAlaGlnThrGluAlaSerProAlaAla 320

1102 CCAGCTTCTTCTCAGCTCTCAATGCGCGGTGTCAGCCCTCGCCGACAGCTGCC 1043
||| ||||| ||||| : : ||| |||
321 ProAsnSerSerAlaAlaProAsnGluProAlaSerGluPro-----ValAla 337
1042 ACAGTCTCTAACTCCAGTGTGTAATGCGCCTCTTCTCAGCTCCAGAGCGCGACT 993
||| ||||| ||||| ||| : : : : : ||| |||
338 GlyArgProGlyThr---AlaAlaSerSerProGluSerAlaSerProValThr 356
982 GCCGCGGTTGAGTCCACCGACA-----CCACTGGCGCGGCC----- 944
||| ||||| : : ||| ||| ||| |||
357 ProArgGlyGluThrProAspThrAlaAlaSerAlaGlyThrProSerAlaGlyArgVal 376
943 ---CCTGCACACAGGTTCCAGGCAATGTCAGCGCGCGTGTAGTCAGGTCAGACCGTGT 887
||| ||||| ||||| ||| : : : : : ||| |||
377 ThrProAlaProAlaProSer-----AlaSerGluGlyAlaSerAlaAlaArgThr 393
886 CTTGGGTTGANTCTCATCTCCATCTCCATATCCGCTCGGCTCCAGGCTCGGTGT 827
||| ||||| ||||| ||| ||||| ||||| |||
394 ProGlyAlaGlySerGlnThrProProIleProAlaThrProIleProAlaThrProAla 413
826 -----ACAGAAAGTGGCGGGAACATGGCGCGCTCCACGCGCTCTCTGTGCA 776
||| ||||| ||||| ||||| ||||| ||||| |||
414 GlyArgSerSerGlyGluSerAlaGlyThrAlaAlaAlaArgProAsnAlaAlaProAla 433
775 -----GGCGGTGGCCAACTGGT---GCAGCGCGC 749
||| ||||| ||||| ||||| ||||| ||||| |||
434 ProValSerGluAspArgSerAspValSerGlyLeuProArgArgGluAspAlaProAla 453
748 CAATCCACA-----CCTAGTAGGCGCGCAGCCAGCCAGCTAGGATCGAAC 698
||| ||||| ||||| ||||| ||||| ||||| |||
454 GluSerSerProValAlaAlaSerProAlaArgGlyAlaSerSerAlaProSerSerAla 473
697 CTGGGTGTCTGCTCCACACGCTCCAGCTCTCT----- 662
||| ||||| ||||| ||||| ||||| ||||| |||
474 ProAlaAlaValProSerArgAlaProValSerGlySerValSerAlaProArg 493
661 -----GCTCCAGTAGCGCG 647
494 ThrAlaProThrAlaProValAlaGluGlnGlyGluValProValSerProSerAlaAla 513
646 GCGCTCGCGCGCGTGTCAATGTTGCTGATGTCGAGATCG---ATCGCCAGAGAAGAC 590
||| ||||| ||||| ||||| ||||| ||||| |||
514 AlaProArgGlyGlyAlaSerSerAlaAlaAlaProSerAlaProAlaAlaArgGly 533
589 GCGCGCACACGAGCATCTGCGCAGCGGCGAGTCTGCTTCTCCAGCTTCGCGCTCGG 530
||| ||||| ||||| ||||| ||||| ||||| |||
534 GlySerGlyAlaAlaGlyGlyAlaAlaGlyGlyAlaSerAlaProAlaAlaAlaArgPro 553
529 CAACGCCACCGCGCTCTGGGACAGGTACGACGAGCTGTCGATGACGTTGTGGA 470
||| ||||| ||||| ||||| ||||| ||||| |||
554 AlaGlnThrProGlyAlaSerAlaGlyAlaSerGlyGlyGly----- 568
469 ACGGTGGAATCATCTGAGCGAGTAGGAGAGTGTATGCGTGGCTCGCTCAGGG 410
||| ||||| ||||| ||||| ||||| ||||| |||
569 -----GluGly 570
409 GCGCAAAATTGGCAAGCGTGGCTCCACGACCTGGAGCATTTCTTCCAGCGCTTCGCTCG 350
||| ||||| ||||| ||||| ||||| ||||| |||
571 ValSer-----AlaArgProSerGlnGlyThrProSerGlyThr 584
349 CTTCTTCTTGGCCA-----CCTCGCAGCAGGAGTGGCAGAGTCCAGCCAGCTAGA 299
||| ||||| ||||| ||||| ||||| ||||| |||
585 ProAlaSerAlaProValAlaAlaGlyArgProAlaGlyGlyGly----- 600
298 TGGAATTGAATCTGGCAGTGCATGTAATCAGCCATCATATGCAATCTTCCAGTGC 239
||| ||||| ||||| ||||| ||||| ||||| |||
601 -----SerGlySerGlyThrSerGlySerGlyAlaProAlaAla 615
238 CACCACCCAGGTCAA-----CCC 221
||| ||||| ||||| ||||| ||||| ||||| |||
616 AlaArgProGlyGlnGlyGlySerGlyAspAlaProValAspLeuGlyLeuThrSerGly 635

QY 220 AGATGAGTTCGAGCGCTCGCGCGCGG----- 191
||| ||||| ||||| ||| |||||
Db 636 ArgGlyGlySerGlyAlaGlyAsnGlyGlyAlaGlyThrGlyGluThrGlyAlaGlyThr 655
190 CAGCAAGGCGCGCGCTCGGCGC-----CGTAGAGCTCTCCAGCGGAGCAG 143
656 ProSerAlaSerArgProGlySerGlyGlyThrGlyAsnGlySerGlyThrGlyGlyGly 675
142 CCGTATCATCGCTCTTCTGCTGCCGCAACCATATGCGCA 101
||| ||||| ||||| ||||| |||||
Db 676 SerGlyThrValProAlaSerGlyArgGlyThrGlyGlySer 689
RESULT 10
S48478
glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast (Saccharomyces cerevisiae)
N;Alternate names: extracellular glucoamylase; mucin-like protein MUC1; protein; YIR019C
C;Species: Saccharomyces cerevisiae
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
R;Accession: S48478; B26877; B26877; S27281; JG6123
R;Rowley, K.
submitted to the EMBL Data Library, October 1994
A;Reference number: S48478
A;Accession: S48478
A;Molecule type: DNA
A;Residues: 1-1367 <ROW>
A;Cross-references: UNIPROT:P08640; UNIPARC:UPI000000056B; GB:Z47047; EMBL:Z38061; NID:96
R;Yamashita, I.; Nakamura, M.; Fukui, S.
J. Bacteriol. 169, 2142-2149, 1987
A;Title: Gene fusion is a possible mechanism underlying the evolution of STA1.
A;Reference number: A91831; MUID:87194600; PMID:3106330
A;Accession: A26877
A;Molecule type: DNA
A;Residues: 1-242 <YAM>
A;Cross-references: UNIPARC:UPI0000168E08; EMBL:M16164; NID:g172522; PIDN:AAA35014.1; PFI
A;Accession: B26877
A;Molecule type: DNA
A;Residues: 762-1331 <YAZ>
A;Cross-references: UNIPARC:UPI0000168E09; EMBL:M16165; NID:g172523; PIDN:AAA35015.1; PFI
R;Pardo, J.M.; Tanez, E.; Zalacain, M.; Claros, M.G.; Jimenez, A.
FEBS Lett. 239, 179-184, 1988
A;Title: Similar short elements in the 5' regions of the STA2 and SGA genes from Sacchar
A;Reference number: S27281; MUID:89031230; PMID:3141213
A;Accession: S27281
A;Molecule type: DNA
A;Residues: 1-31 <PAR>
A;Cross-references: UNIPARC:UPI0000168E60; EMBL:X13857; NID:g4551; PIDN:CAA32069.1; PID:5
R;Lambrechts, M.G.; Bauer, F.F.; Marmur, J.; Pretorius, I.S.
Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996
A;Title: Muc1, a mucin-like protein that is regulated by Msi10, is critical for pseudohy
A;Reference number: JG6123; MUID:96323237; PMID:8710886
A;Accession: JG6123
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-1367 <LAM>
A;Cross-references: UNIPARC:UPI000000056E; GB:U30626; NID:g1304386; PIDN:AAC49609.1; PID
C;Genetics:
A;Gene: SGD:MUC1; STA2; MAL5; DEX2; SGD:S0001458
A;Cross-references: MIPS:YIR019c; SGD:S0001458
A;Map position: 9R
C;Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase
C;Keywords: glycosidase; hydrolase; polysaccharide degradation; transmembrane protein
F;5-21/Domain: transmembrane #status predicted <TM1>
F;1350-1366/Domain: transmembrane #status predicted <TM2>
Alignment Scores:
Pred. No.: 5.26e-07 Length: 1367
Score: 257.00 Matches: 170
Percent Similarity: 38.3% Conservative: 85
Best Local Similarity: 25.6% Mismatches: 298
Query Match: 7.1% Indels: 112
DB: 1 Gaps: 24
US-10-620-914-44 (1-1947) x S48478 (1-1367)

A;Reference number: Z22899; MUID:97166151; PMID:9013550
A;Accession: T45025
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-3570 <DES>
A;Cross-references: UNIPARC:UPI000016AD88; EMBL:Z72496; MID:g1834502; PIDN:CAA96577.1; P
A;Experimental source: placenta
C;Genetics:
A;Gene: MUC5B

Alignment Scores:
Pred. No.: 1,07e-06 Length: 3570
Score: 253.00 Matches: 185
Percent Similarity: 35.0% Conservative: 67
Best Local Similarity: 25.7% Mismatches: 291
Query Match: 7.0% Indels: 176
DB: 2 Gaps: 35

US-10-620-914-44 (1-1947) x T45025 (1-3570)

QY	17	ACGGCCGCGCTGGAGCTACACCAAGAAGAACTTCTCCCTGGAGAAGCTCAAGCTCAGCA	76
DB	2879	SerThrGlySerThrAlaThrProSerSerThrProGly-----	2891
QY	77	GCATGAAGATGACCTGACCGTTCTGGCCCATATGTGTTCCGACGACGAAGGGCGATG	136
DB	2892	-----ThrAlaProProProLysValLeuThrSerProAlaThrThrProThr	2907
QY	137	ATCACGCTGCTCGCTGGAGAGCTTACGGGCCCCAGGCGCGTGTCTGTCGCGCC	196
DB	2908	AlaThrSerSerLysAlaThrSerSerSerProArgThrAlaThrLeuProVal	2927
QY	197	TGCGCCGAGCGCTCGAAGCTCATCTGGGTTGACCTGGGTGGTGGCACTGGGAGAAATGTCG	256
DB	2928	LeuThrSerThrAlaThrLysSerThrAlaThrSerValThrProile-----	2943
QY	257	ATATGATGGCTGATTACATCGACTGGCGGAAGTTCAAGTCCATCTACGTGGTGCACCTGT	316
DB	2944	-----ProSerSerThrLeuGlyThrThr	2951
QY	317	GCCACTCGCTGTGCGAGTGGCCCAAGAAGGAAGGCAAGGCGAAGGCTGGAAGATGTC	376
DB	2952	GlyThr-----LeuProGluGlnThrThrProValAla-----	2963
QY	377	AGTCTGCGAGGCGCGCTGTCGAATTTGCGCCCTCGAGGCGACCGGACGCTCATCA	436
DB	2964	-----ThrMetSerThrIleHisProSerSerThrProGluThrHis	2978
QY	437	CCTTCTCTACTCGCTCAGCATGATCCACCGTTCACACAGCTATCGACGAGCTTGCT	496
DB	2979	ThrSerThrValLeuThrLysAlaThrThrArgAlaThrSerSerThrPro	2998
QY	497	CGTACCTGTCACCAAGCGCTGGTGGCGTTCGCGCTCTACGTGAGCGGCAAGTACG	556
DB	2999	SerSerThrProGlyThrThr-----TrpIleLeuThrGluLeuThrAlaAlaThrThr	3017
QY	557	ACCTGC-----CCCTGCGCC-----AGATGCGCTCGTGGCGCGCTTCTCT	598
DB	3018	ThrAlaGlyThrGlyProThrAlaThrProSerSerThrProGlyThrThrTrpIleLeu	3037
QY	599	GGCGATCGATTTTCGACATGCAACATTTGACATTCGCGCCCGGAGCGCGCTTCTCT	658
DB	3038	ThrGluLeuThrThrThrAlaThrThrThrAlaSerThrGlySerThrAlaThrLeu---	3056
QY	659	AGCAGAAGCTGGAGCGGTGGGAGCAGACACCCAGGTTTCATCCCTAGTCCCGT	718
DB	3057	-----SerSerThrPro-----GlyThrThrTrpIle	3065
QY	719	GGCTGGCGCGCCCTACTACGTGTGATTTGGCGCGCTGCGCCA-----GGTGGCGCGCC	775
DB	3066	LeuThrGluProSerThrThrAlaThrValThrAlaProProGlySerThrAlaThrAla	3085
QY	776	TGCACGAGGAGCGGTGGAGCGCGCGCCA-----TGTTCCCGCCCACTTCC	823

DB	3086	SerSerThrGlnAlaThrAlaGlyThrProHisValSerThrThrAlaThrThrProThr	3105
QY	824	TGTACACGCGAGTGTGGAGGAGCCCGAGCGGATATGGAGGTGGATGAGATCAACCCCA	883
DB	3106	ValThrSerSerLysAla-----ThrProSer-----SerSerPro	3117
QY	884	AGGACACGCTGTGACCTGACTAGCGCGGCTGCAATGCCCTGAAACCTGCTGGTGCGG	943
DB	3118	GlyThr-----AlaThrAlaLeuPro-----	3124
QY	944	GGGCGCGCAGGTGGTGTGGTGGACTGCA-----ACCCGCGCAGTGGCGCTTCTGAGC	1000
DB	3125	-----AlaLeuArgSerThrAlaThrProThrAlaThrSerPheThrAla	3140
QY	1001	TGAAGAAGGTGGCCATTTCAGCAGC-----TGGAGTTTGGAGAGCTGGGCGC	1048
DB	3141	-----IleProSerSerSerLeuGlyThrThrTrpThrArgLeuSerGlnThrThr	3157
QY	1049	TGTTGCGGAGGCGGTGCACCCGCGCATTTGAGGAGCTGTACGAGAGAAGCTGGCGCCT	1108
DB	3158	ThrProThrAlaThrMetSerThrAlaThrProSerSerThrProGluThrValHisThr	3177
QY	1109	TCC-----TGTGCAACACGACCACTTCTGGTCCAGCGCC-----	1147
DB	3178	SerThrValLeuThrThrThrAlaThrThrThrGlyAlaThrGlySerValAlaThrPro	3197
QY	1148	TCTGTGTACTTCCAGCAGCGCTGTACTTACCAGGCGGCATGGGCAAGCTGTGGGTGC	1207
DB	3198	SerSerThrProGlyThrAlaHisThrThrLysValProThrThrThrThrGlyPhe	3217
QY	1208	TGCAGTGCCTGGCGGTGGTGGTGGGACTGGGCAAGCCGTCAGCGCTCGCCA-----	1261
DB	3218	Thr-----AlaThrProSerSer-----SerProGlyThr	3227
QY	1262	-----ACGGCGCCCAATGGAGGAGCAGCGCGCTGTGGGACAGCAACATGCTCATCC	1315
DB	3228	AlaLeuThrProProValTrpIleSerThrThrThrThrProThrThrThrProThr	3247
QY	1316	ACTTCTGTGAAGAAGC-----GGCCCAAGCGCTGGTGGTGTTCGTCACAGTTCGTGAGCC	1372
DB	3248	ThrSerGlySerThrValThrProSerSerIleProGlyThrThrHisThrAlaArgVal	3267
QY	1373	TGTTGCTCTTCAACAGCGCTGTGTGGTGGCGCGCGCTGGCGGCGGAGCAGTACG	1432
DB	3268	LeuThrThrThrThrValAlaThrGlySerMetAlaThrProSerSerSerThr	3287
QY	1433	CGTGTATCAAGGCGGAGCGCATCCCATTTGAGNACTTACTTACTACAACTGCTCACC	1492
DB	3288	GlnThrSer-----GlyThrProProSerLeuThrThrThrAlaThrThrIleThrAla	3305
QY	1493	TGGCGGAGAACTCGCACGTGGCGCAAGCAGAACTTCTTACTACAACTGCTCACC	1552
DB	3306	ThrGlySerThr-----ThrAsnProSerSerThrProGlyThrThrProleProVal	3324
QY	1553	-----AGTTCTTGGCGCAAACTGCTGCCACCTTCTGCGCGAGCGGCTTCCGCCACCTCA	1609
DB	3325	LeuThrSerMetAlaThrThrProAlaAlaThrSerSerLysAlaThrSerSerSer	3344
QY	1610	-----AGAGTGGGTGGGCAACCTGACCGTCTCCACCAACTTCTTCATGGAGAGCTCA	1666
DB	3345	ProArgThrAla-----ThrThrLeuProValLeuThrSerThrAlaThrLysThr	3362
QY	1667	AAGCGCGCA-----CCTACACCAAGGTGATTTCTGATGACCAACGCTGGCTGGATA	1720
DB	3363	AlaThrSerPheThrThrProleProSerSerThrLeuThrThrThrThr	3379
QY	1721	TGCGCGTGGCCCAACGAGCTGGCGAGTGCCTGGCCAAAGCAGGTGGCGCGCGCG	1774
DB	3380	ValProAlaGlnThrThrThrProMetSerThrMetSerThrIleHisThrSerThr	3399
QY	1775	-----GCATCG	1780

Db 3400 ProGluThrThrHisThrSerThrValLeuThrThrAlaThrMetThrAlaThr 3419
QY 1781 TCATCTGGCGCTCGCTCCTCAGCC-----CGC 1810
Db 3420 AsnSerThrAlaThrProSerThrLeuGlyThrThrArgIleLeuThrGluLeuThr 3439
QY 1811 CCTACGCGGAGCTGATCCAGAGCGGCTTCACGCTGCGCTGCATCCGCGCGCCACTC 1870
Db 3440 ThrThrAlaThrThrAlaAlaThrGlySerThrAlaThrLeuSerSerThrPro--- 3458
QY 1871 AGGGCTACATGG-----ACCGC 1888
Db 3459 GlyThrThrTrpIleLeuThrGluProSerThrIleAlaThrValMetValProThrGly 3478
QY 1889 TCAACATGTACAGCTCTTACATGCGCGCGGAGGCGGCAAGAGGCAACT 1945
Db 3479 SerThrAlaThrThrSerSerThrLeuGlyThrAlaHisThrProLysValValThr 3497
RESULT 12
T45463
membrane glycoprotein [imported] - equine herpesvirus 1
C;Species: equine herpesvirus 1
C;Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 09-Jul-2004
C;Accession: T45463
R;Kirisawa, R.; Kobayashi, T.; Kawakami, Y.; Iwai, H.
J. Equine Sci. 7, 79-87, 1996
A;Title: Nucleotide sequences of open reading frames 1, 24 and 71 of an attenuated equine herpesvirus 1 glycoprotein homolog
A;Reference number: 222973
A;Accession: T45463
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-867 <KIR>
A;Cross-references: UNIPROT:O39782; UNIPARC:UPI000000EF109; EMBL:D88734; PIDN:BAR20038.1
A;Experimental source: isolate 3F clone; strain BK343
C;Genetics:
A;Note: ORF71
C;Superfamily: equine herpesvirus glycoprotein X; equine herpesvirus 1 glycoprotein homolog
Alignment Scores:
Pred. No.: 1.06e-06 Length: 867
Score: 251.00 Matches: 119
Percent Similarity: 41.0% Conservative: 81
Best Local Similarity: 24.4% Mismatches: 208
Query Match: 6.9% Indels: 80
DB: 2 Gaps: 14
US-10-620-914-44 (1-1947) x T45463 (1-867)
QY 473 ACAAGTCATCGACAGGCTTGCTGCTGCTCCCAAGAGCGGCTGGTGGGCGTTGCCG 532
Db 46 ThrThrAsnSerSerSerSerProThrThrThrSerProThrThrSerSerProPro 65
QY 533 ACTTCTACGTAGCGGCAAGTACAGCTGCCCTGCCGCCAGATGCCCTGTGCGCGCGTT 592
Db 66 ThrSerThrHisThrSerProSerSerThrSerThrGlnSerSerThrAlaAla 85
QY 593 TCTTCTGGGATCGATCTTCGACATCGACACATTGATCGCGCGCGCGCGCGCT 652
Db 86 ThrSerSerAlaProSerThrAlaSerSerThrThrSerIleProThrSerThr 105
QY 653 ACCTGGAGCAGAGCTGGAGCGGTGGGAGCAGACACCCAGGGTTCGATCCCTACG 712
Db 106 ThrGluThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 121
QY 713 TGCGGTGGCTCG 772
Db 122 ThrThrThrAlaAlaProThrThrAlaAlaThrThrThrThrThrThrThrThr 141
QY 773 CCCTGCAGGAGCGGCGGTGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 832
Db 142 ThrSerAlaGluThrThrThrThrThrThrThrThrThrThrThrThrThrThr 160
QY 833 AGTCGTGGAGGACCCCGAGCGGATATGGAGGTGATGGAGATCAACCCCAAGGACACGG 892

161 -----ThrProThr-----SerThrThrThrThr----- 168
QY 893 TGTGACCTGACTAGCGCGGCTCAATGCCCTGAACCTGCTGTGTCAGGGGCGCGCC 952
Db 169 -----ThrAlaThrThrValProThrThr----- 177
QY 953 AGGTGTGTGGTGGACTGCAACCCCGCAGTCGGCGCTTCTGGAGCTGAAGAAGTGG 1012
Db 177 ----- 177
QY 1013 CCATTGACAGCTGGAGTTTGAGGACGTGTGGCAGCTGTTGGCGAGGGGTGCACCCGC 1072
Db 178 -----AlaSerThrThrThrThrAlaAlaThrThrAlaAlaThrThrAlaAlaThrThr 195
QY 1073 GCATTGAGAGCTGTACGAGAGAGAGTGGCGGCTTCTGTGCGCAACCCAGCCACAAC 1132
Db 196 ThrAlaAlaThrThrThrAlaAlaThrThrAlaAlaThrThrAlaAlaThrThr 215
QY 1133 TCTGGTCCAAAGCGCTCTGCTACTTCCAGCAGCGCTGTACTACAGGCGCGCATGGCA 1192
Db 216 ThrAlaAlaThrThrThrAlaAlaThrThrAlaAlaThrThrAlaAlaThrThr 235
QY 1193 AGCTGTGCTGGTGTGTCAGTGCCTGCGCGCTGTGTGGAGCTGGGCAAGACCGTCAAGC 1252
Db 236 ThrAlaAlaThrThrThrAlaAla-----ThrThrThr 246
QY 1253 GCCTCCCAAGCGCGCCCAATGGAGAGCAGCGCGCTGTGTGGAGCAGCAACATGCTCA 1312
Db 247 AlaAlaThrThrThrAlaAlaThrThrSerSerAlaThrThrAlaAlaThrThrAla 266
QY 1313 TCCACTTCGTGAAGAACCGCGCCCAAGCGCTGTGTGGTGTTCGTCGAAGTGTGTGAGCC 1372
Db 267 AlaThrThr-----ThrAlaAlaThrThrThrAlaAlaThrThrAlaAlaThrThr 284
QY 1373 TGTGTCTTCAACAAGCGCTGCTGTGTTCGGCGCGCGGTGGCGGCAAGCAGTACG 1432
Db 285 ThrAlaAlaThrThrThrAlaAlaThrThrAlaAlaThrThrAlaAlaThrThr 304
QY 1433 CGCTGATCAAGCGGAGCGCA-----TCCCATTTGAGACTACATCGCGCGCACCATGG 1486
Db 305 ThrAlaAlaThrThrThrAlaAlaThrThrAlaAlaThrThrAlaAlaThrThr 324
QY 1487 ACGCGTGGCGGAGAACTCGCAGCTGCGCAAGCAGAACTACTTCTACTACA----- 1537
Db 325 ThrAlaAlaThrThrThrGlySerProThrSerGlySerThrSerThrGlyAlaSer 344
QY 1538 ACTGCTCACCGCAAGTCTCTCGCGGCAACTGCGCCCACTACTCTCGCGGCGCGCT 1597
Db 345 ThrSerThrProSerAlaSerThrAlaThrSerAlaThrProthr----- 359
QY 1598 TCGCCACCTCAAGATGGGTGGTGGAGCAACCTGACCGCTCCACCACT-----TCT 1651
Db 360 SerThrSerThrSerAlaAlaAlaThrThrSerThrProThrProThrSerAlaAlaThr 379
QY 1652 TCATGGAGGAGCTCAAGCGCGCACCTTACCAAGGTGATTCGATGGACACCGTGGACT 1711
Db 380 SerAlaGluSerThrGluAlaProThrSer-----ThrProThrThr 394
QY 1712 GGCTGGATATGCCGTGGCGCAAGAGCTGGCGGAGTCCCTGGCGCAAGCAGGTTCGCGCG 1771
Db 395 AspThrThrThrProSerGluAlaThrThrAlaThrThrSerProGluSerThrThrVal 414
QY 1772 CGCGCATCTCATCTGGCGCTCGCCCTCCCTGACCGCGCGCTAGCGAGCTGATCCAGA 1831
Db 415 SerAlaSerThrSerAla-----ThrThrThrAlaPheThrThrGluSerHisThrSer 433
QY 1832 AGCGCGCTTCGAGCTGCGCTGCATCCGCGCGCGCACTCAGGCGTACATGACCGCGCTCA 1891
Db 434 ProAspSerSerThrGlySerThrSerThrAlaGluProSerSerThrPheThrLeuThr 453
QY 1892 ACATGTACAGT---CCTTCTACA 1912

783	Qy	GGAGCGCGTGGACGGCC	-----	800
761	Db	GIyAlaLeuGIyLeuGIyAlaGIuLeuValGIyGlnArgValValGIuHis	780	
801	Qy	---GCCCATGTT---	CCCCGCCACCTTCGTACACGCACGTCGTGGAGACCCCGA	851
781	Db	HisAlaHisValLeuGIyValGIyTyrLeuPro---	HisProGIyAlaAlaAlaGIu	799
852	Qy	CGCGCATATGGAGGTGATCGAGATCAACCCCAAGACACGCGTCTCACCTGCTAGCGG	911	
800	Db	ArgGIyAlaAlaAlaArgGIyAspValArgGlnGIyArgValGIuArgArg	819	
912	Qy	CGG-----	CTGCATTCCTGAACTCTGCTGTCAGGGGGCCGCCA	953
820	Db	AlaProGluPheGIyGluAspLeuLeuValHisGluGIyAlaGIyHisLeuGIyArgAla	839	
954	Qy	---GGTGGTCTCGGTGGAGTG---	CAACCCCGCGCAGTCGGCGCTTCGTGGAGTGAAGAA	1007
840	Db	ValGIyGIyGluGIyArgLeuGIyGIyProArgArgValGIyLeuAlaGIyArgAsp---	858	
1008	Qy	GGTGGCCATTACGACGCTGGAGTTTCAGGACGTGTGGCAGCTCTTCGGCGAGGG---	1061	
859	Db	-----	AlaAlaGluAlaAlaValGIyArgGIyValLeu	869
1062	Qy	-----CGTGCACCC	-----	1070
870	Db	GIyHisGIyProGluArgAlaProGluProValValLeuGIyGIyGIyGIyGIyGly	889	
1071	Qy	---GCCANTGAGGACTGTACA---	GAAGACGTGGCGCCCTTCCTGTGCGCA	1118
890	Db	GlnGlnArgGIySerGIyValArgSerGIyProGluSerGIyAlaAlaLeuAlaPro	909	
1119	Qy	AACGACGCCAACCTTCGTGCCAAGCGCTCTGGTA---	CTTCGACGACGCGCTGTACTA	1175
910	Db	GlyProProValLeuPheValValAlaValAlaValAlaValProAlaGlu	926	
1176	Qy	CCAGGCGCGCATGGCAA-----	GCTGTGCTGGGTGCTGCAGTGCCTGCG	1220
927	Db	---GlyArgAlaGIyGluProLeuValLeuLeuAlaValProGIyAlaAlaGIyProGly	945	
1221	Qy	CGT-----	1223	
946	Db	ArgAlaAlaLeuLeuAlaProLeuGIyArgTrpValArgAlaGIyGIyGIyAla	965	
1224	Qy	-----	GGTGTGGCAGCTGGCGAAGACCGCTCAAGCGCCT	1256
966	Db	GlyValAlaGIyAlaGIyGluAlaGIyLeuGIyAlaGIyAlaGIyLeuGIyAlaGIy	985	
1257	Qy	CGCAACGCGCCACCAATGGAGGACGAGCGCGCTGTCTGGGACAGCAACATGCTCATCCA	1316	
986	Db	AlaGIyLeuGIyAlaGIyGlyAlaGIyGlyProGIyAlaGIyGluAlaGIyGIyAla	1005	
1317	Qy	CTTGTGTGAACCGGCCCAA	-----GCCGTGCT	1346
1006	Db	ArgArgArgArgArgArgTrpAspAspGluAlaGIyLeuLeuGIyProGluArgGly	1025	
1347	Qy	GTGGCTTTCGTCACAGTTCTGTGACGCTGTGCTCTT-----	CAACAAGCGCCT	1394
1026	Db	GlnAlaGIyArgGIyLeuArgGlyProGIyProArgGIyGIyLeuGIyGluProGIyArg	1045	
1395	Qy	GCTGTGTTTCGGCGGGGTGCGCGGCAAGCAGTACGGCGCTCATCAAGCGCAGCGCAT	1454	
1046	Db	GlyHisValGIyArgGIyGluGluGIyArgGIyVal-----	GlyProGIyGIyLeuAla	1063
1455	Qy	CCCCATTGAACTACATCGCGCGCACCATGAGCGCGTGGCGGA	-----	1499
1064	Db	GIyAlaGIyProValHisAlaValAlaHisGlnArgArgHisGIyAlaGIyAspGIyGly	1083	
1500	Qy	-----	GNACTCGCACGTGGCGACAGCAACTCTTCTACTACAA	1538
1084	Db	AspArgValArgGIyLeuProProLeuGIyArgAlaGIy-----	1096	
1539	Qy	CTSCCTCACCGGCAAGTCTCTGGCGACAACTG-----	CCCCCACTTACTCTGCGCGAGCG	1592

Db 454 ProSerThralaThrProSerThr 461

RESULT 13
B45344
probable nuclear antigen - suid herpesvirus 1 (strain Kaplan)
C;Species: suid herpesvirus 1
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 09-Jul-2004
C;Accession: B45344
R;Vicek, C.; Kozmik, Z.; Paces, V.; Schirm, S.; Schwyzer, M.
Virolgy 179, 365-377, 1990
A;Title: Pseudorabies virus immediate-early gene overlaps with an oppositely oriented orf
A;Reference number: A45344; MUID:91021039; PMID:2171211
A;Accession: B45344
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-1733 <VIC>
A;Cross-references: UNIPROT:P33485; UNIPARC:UPI0000138ADE; GB:M34651; NID:G334070; PIDN:
C;Superfamily: pseudorabies virus 1 nuclear antigen

Alignment Scores:

Pred. No.:	1.47e-06	Length:	1733
Score:	249.50	Matches:	201
Percent Similarity:	31.7%	Conservative:	32
Best Local Similarity:	27.3%	Mismatches:	257
Query Match:	6.9%	Indels:	245
DB:	1	Gaps:	40

US-10-620-914-44 (1-1947) x B45344 (1-1733)

Qy 282 GCGAAGTTCACGTCCATCTACGTGFCGACCTGTGCCACTCGCTGTGGAGTGCCAA 341
||| ||| : : : ||| ||| ||| : :
Db 564 GlyHisValLeuAlaAlaValHisGlyGlyGluHisAlaLeuAlaValGlyAlaArg 583
||| ||| : : : ||| ||| ||| : :
Qy 342 GAAGAAGCGCAAGCCAGGCTGGAAGATGTCACAGTGCAGTGTGGAGCCGACGTTGCCA 401
: : : ||| ||| ||| ||| ||| ||| : :
Db 584 GluArgAspArgGlyProGly-----ProGlyAlaGly----- 594
||| ||| ||| ||| ||| ||| ||| : :
Qy 402 ATTTGGCCCCCTGAGGCGACCGCGAGCTCATCACCTCTCTCTACTC----- 449
||||| ||| ||| ||| ||| : : : ||| : :
Db 595 -----AlaHisArgValAlaHisValValLeuAlaAlaGluAlaGlnArg 609
||| ||| ||| ||| ||| ||| ||| : :
Qy 450 -----GCT 452
||| ||| ||| ||| ||| ||| ||| : :
Db 610 LeuGlyProGlyValGlnAlaGlyGluGlyGlyLeuHisAlaGlyGluAlaGlyArgAla 629
||| ||| ||| ||| ||| ||| ||| ||| : :
Qy 453 CACGATATCCACCCTTCACAACGTCATCGACACGAGTGTCTGTACCTGTCCCAAGA 512
||||| : : : ||| ||| ||| ||| ||| : :
Db 630 HisAspGlyAlaargVal---GlnHisGlyArgAlaGluLeuAlaalaProGlyProAla 648
||| ||| ||| ||| ||| ||| ||| : :
Qy 513 CGGCT-----GTTGGCGTTCGCCACTTCTACGTGAGCGCAAGTACGACCTGCCCT 566
||| ||| ||| ||| ||| ||| ||| : : : ||| : :
Db 649 HisGlyAlaLeuGlyGlyArgValGlnAlaAspValaspVal---ValValPro 667
||| ||| ||| ||| ||| ||| ||| : : : ||| : :
Qy 567 GCGCCAGATGCCCTGTGTGCGCGCTTTCTTCTGGCGATCGATCTTCGACATCGACACAT 626
||| ||| ||| ||| ||| ||| ||| : : : ||| : :
Db 668 HisGlyArgAlaValargGlyProValLeuaspGlyVal-----GlnHis 682
||| ||| ||| ||| ||| ||| ||| : : : ||| : :
Qy 627 TGACAT-----CGGCCCGAGCGCGCGCTTACTGTGAGCA----GAA 665
||||| ||| ||| ||| ||| ||| ||| : : : ||| : :
Db 683 AspGluProAlaProArgAlaGluProArgAlaGluValLeuHisGlyAlaGlyGlu 702
||| ||| ||| ||| ||| ||| ||| : : : ||| : :
Qy 666 GCT-----GGAGCGGTGGGA 683
||| ||| ||| ||| ||| ||| ||| : : : ||| : :
Db 703 AlaGluValProArgArgGluGlnGlnHisProLeuGlyValGluAlaAlaaspValGly 722
||| ||| ||| ||| ||| ||| ||| : : : ||| : :
Qy 684 GCAGAACACCCAGGTTTCGATCCCTACGTGCGCTGGCTGCGCGCGCCCTTACTACGTGTG 743
||| ||| ||| ||| ||| ||| ||| : : : ||| : :
Db 723 Ala-----ProGlyProValProGlyProGlyValargValargAlaGluAlaVal 740
||| ||| ||| ||| ||| ||| ||| : : : ||| : :
Qy 744 GAT-----TGCGCGCTGCCAGCGTTGGCCAGCCCTGTGCACGA 782
||| ||| ||| ||| ||| ||| ||| : : : ||| : :
Db 741 GlyGluGlyGluGlnArgArgGluAlaAlaAlaAlaargValProGlyArgAlaArg 760
||| ||| ||| ||| ||| ||| ||| : : : ||| : :

```

Db      1097  ---ProGlyAspArgValAlaGluArgGluGlnArgGlyArgHisLeuLeuGluAlaGly 1115
QY      1593  GGCCTTCGCCACCTCAAGAGTGGCGTGGT-----GGACAACTGACCGTCTC 1640
Db      1116  Gly-----ProGluGlyGlyArgGlyAlaGlyArgGlyGlnProGluArgAla 1132
QY      1641  ---CACCAACTTCTTCATGAGGAGCTCAAGCGGCACCTACACCAAGGTGATTCTGAT 1697
Db      1133  GlyGlnGlnAlaLeuGluAspAlaA-----AlaGlyGlnAspAla 1146
QY      1698  GGACCACTGAGTGGCTGGATATGCCCGTGGCCAAACGAGCTGGCGAGTGCCTGGCCAA 1757
Db      1147  GlyValArgGlnLeuAlaGlyHisAlaAlaGlyLeuArgGlyGlyGluGlyAlaAsp 1166
QY      1758  GCAGTTGGCCGGCGGCGATCGTCATCTGGCGTCCGCTCCCTCAGCCCGCCCTACGC 1817
Db      1167  AlaGlyAlaGluGly-----LeuAspGlyArgLeuProGlyAlaGlyValArg 1182
QY      1818  CGAGCTGATCCAGAAGGC-----GGG 1838
Db      1183  GlyAlaAlaArgValGlyHisValGlyValGlyProAlaGluAlaLeuGlnAspGluGly 1202
QY      1839  CTTTCAGCTGGCTGCATCGCGCGCCACTCA---GGGCTACATGGACCGCGTCAACAT 1895
Db      1203  LeuLeuGlyAlaIleValAlaAlaAlaHisGlyHisGlyAlaHisArgValArgGlnGly 1222
QY      1896  GTACAGCTCCTCTTA-----CATGGCCCGCGGAAGGGCGCCNA 1934
Db      1223  ProGluArgValLeuGlyGlyHisGlyValProAspValArgGln 1237

RESULT 14
QB83
BHLF1 protein - human herpesvirus 4 (strain B95-8)
C;Species: human herpesvirus 4, Epstein-Barr virus
C;Date: 25-Feb-1985 #sequence_revision 25-Feb-1985 #text_change 09-Jul-2004
C;Accession: A03742
R;Bankier, A.T.; Deininger, P.L.; Farrell, P.J.; Barrell, B.G.
Mol. Biol. Med. 1, 21-45, 1983
A;Title: Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8 Epstein-Barr virus
A;Reference number: A93065; MUID:85035713; PMID:6092825
A;Accession: A03742
A;Molecule type: DNA
A;Residues: 1-660 <BAN>
A;Cross-references: UNIPROT:P03181; UNIPARC:UPI0000000CDC0
R;Baer, R.; Bankier, A.T.; Biggin, M.D.; Deininger, P.L.; Farrell, P.J.; Gibson, T.J.; H
Nature 310, 207-211, 1984
A;Title: DNA sequence and expression of the B95-8 Epstein-Barr virus genome.
A;Reference number: A03794; MUID:84270667; PMID:6087149
A;Contents: annotation; protein coding region
C;Comment: The sequence contains four perfect repeats (residues 149-273, 274-398, 399-52
C;Superfamily: human herpesvirus 4 BHLF1 protein

Alignment Scores:
Pred. No.: 1.93e-06 Length: 660
Score: 246.00 Matches: 206
Percent Similarity: 32.7% Conservative: 29
Best Local Similarity: 28.7% Mismatches: 257
Query Match: 6.8% Indels: 228
DB: 1 Gaps: 42

US-10-620-914-44 (1-1947) x QB83 (1-660)
QY      155  AGACTTCTACGGCCCGCCAGG-----CGCTGCCTTTCTCTGCC 193
Db      6   GlnSerAlaArgGlyProArgThrProLeuProHisCysProProCysLeuPro 25
QY      194  GCCTGGCGCGCTCGAACCTCATCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 253
Db      26  GlyAlaProAspGlnGlnThrArgArgLeuPro-ProGlyTrpGly----- 40
QY      254  TCGATATGATGGCTGATTACATCGACCTGGCGAAGTTCAAGTCCATCTACGTGTCGACC 313

```

```

Db      41  -----GlnArgThrAlaProThrGlnValGlyLeuAlaAspAla 54
QY      314  TGTGCCACTCGCTGTCGA-----GGTGCCCAAGAAGA 346
Db      54  aSerProAspGluLeuGlnAspGlnAlaSerGlyAlaArgProGlyGlyGlyAsnArgVa 74
QY      347  AGCGGAAGGCCAAGGGCTG-----GAAGAATGTCAGGTCTGGAGGCGCGCTTGCC 400
Db      74  lGlyAlaGlyArgGlyArgProGlyThrProAlaProSerArgGlnSerArgArg----- 92
QY      401  AATTTCGGCCCTGAGGGCAGCGGACGCTCATCCTCTCTCTCTCTCTCTCTCTCTCTCTCT 460
Db      93  ---ThrGlyPro-----AlaGluGlnAlaAspHis-----AlaHisSerAs 105
QY      461  TTCCACC-----GTTCCCAACGTCAT-----CGACAGGCTTGTCTGT 499
Db      105  nProThrGlyGlyCysSerAspProGlnArgSerProArgThrArg-GlnAla---GlyT 124
QY      500  ACCTGTCCCAAGACGGCCTGGTGGCGTTCCTCCGACTTCTACGTGAGCGGCAAGTACGACC 559
Db      124  yAlaLeuGlyGluGlySerAlaGlyLeuGlySer-----ArgGlyProArgProH 141
QY      560  TGCCCTTCGCCAGATGCCCTGTGCGCGCT-----TTCTTTCGGCGAT 604
Db      141  isProAlaPheGlnValGlnTrpSerAlaArgAsnProGlyCysProArgThrTrpArg- 160
QY      605  CGATCTTCGACATCGCAACATTTGACAT----- 632
Db      161  ArgArgSerGlyAla-GlnArgGlyHisProProProGlyAlaGlyGlnArgProSerGl 180
QY      633  -----CGGCCCGAGCGCGCCCTTCTTCTGGAGCAGAAAGCT----- 668
Db      180  yProThrGlyGlyArgProAlaAlaProGlyAlaProGlyThrProAlaAlaProGlyPr 200
QY      669  -----GGAGCCCGTGTG-----GGAGCAGACACCCA-----GGTTCGATCC 706
Db      200  oGlyGlyGlyAlaAlaValProSerGlyAlaThrProHisProGluArgGlySerGlyPr 220
QY      707  CCTACGTGCGCTGGCTGCG-----CGCCCCCTACTACGTGT 742
Db      220  oAlaAspProProAlaAlaAlaAlaArgLeuProProGluArgGlnGluProArgLeuProGl 240
QY      743  GGATTCGCGCTCCCGCGCTTGGCCAGC-----CCTGCACGAGGAGC---GGGTG 792
Db      240  nAspLeuAlaAlaAlaGlnArgCysProAlaGlyProProProThrArgSerGlyAlaAl 260
QY      793  GAGCGCGCGCCATGTTCCCGCCACTTCTCTACGCGAGTCTGGGAGGAGCCCGAG 852
Db      260  aAlaGlnArgThrHisArgArgProProGlyCysProArgSer---AlaArgAsnProGl 279
QY      853  C-----CGCATATGGAGGTGATGGAGATCAACCCCAAGGACACGCTGCTGACCTGACT 906
Db      279  yCysProArgThrTrpArg-----ArgArgSerGlyAlaGlnArg-GlyHisPro--- 295
QY      907  AGCGGCGCTGCAATGCCCTGAACCTGCTGGTCAGGGGCGCGCCAGAGTGTGTCGGTG 966
Db      296  -----ProProGlyAlaGlyGlnArgProSerGlyProThrG 308
QY      967  GACTGCAACCCCGCGAGTGGCGCTTCTGGA-----GCTGAAGAGGTGGCCATT 1017
Db      308  lyGlyArgProAlaAlaProGlyAlaProGlyThrProAlaAlaProGlyProGlyGlyG 328
QY      1018  CAGCAGCTGGAGTTGAGGACGTGTGGCAGCTGTCGGCGAGGCGGTGCACCCCGCATT 1077.
Db      328  lyAlaAlaValProSerGlyAlaThrProHisProGluArgGlySerGlyProAlaAspP 348
QY      1078  GAGGAGCTGTACGAGAGAAGAGTGGCCCTTCTCTGTCGAAACACGACCACTTCTGG 1137
Db      348  roProAla-----AlaAlaArgLeuProProGluArgGlnGluProArgLeuP 364
QY      1138  TCNA-----GGCCCTCTGTACTTCCAGCAGCGCTGTACTACCGGGGGCGCATG 1188
Db      364  roGlnAspLeuAlaAlaGlnArgCysProAlaGlyProPro---ProThrArgSerG 383

```

QY 1189 GCGAAGCTGTCTG-----GGTGTGCAAGTCCCTGGCGGTGGT 1227
Db 383 lYAlaAlaAlaGlnArgThrHisArgArgProProGlyCysProArgSerAlaArgAsp 403
QY 1228 CTGGGACTGGGCAAGAC-----CGTCAAGGCTCCGCAACGGCCACCAATGGAGGAG 1281
Db 403 roGlyCysProArgThrTrpArgArgArgSerGlyAlaGlnArgGlyHis-----P 420
QY 1282 CAGGCGCTGTGGGACAGCAACATCTCATCCATTCGTGAAGAACGGGCCCAAGCCG 1341
Db 420 roProProGlyAlaGlyGlnArgProSerGlyProThrGlyGlyArgProAlaAlaProG 440
QY 1342 CTGGTGTGGCTGTTCGTCAAGTTCGTGAGCTGTGTCTTCAACAGGCGGTCTGTGG 1401
Db 440 lYAlaProGlyThrProAlaAlaProGlyProGly----- 451
QY 1402 TTCGGCGGCGGTGCGGGCAAGCAGTACGCTGATCAAGGGGAGCGGATCCCATTT 1461
Db 452 -----GlyGlyAlaAlaVal-----ProSerGlyAlaThrProHisP 464
QY 1462 GAGAACTACATCGCGCGCACCATGACGGCGTGGCGGAGAACTCGCACGTGGCGAAGCAG 1521
Db 464 roGlu-----ArgGlySerGlyProAlaAspProAla - 475
QY 1522 AACTACTTCTACTACAACTGCTTCAACCGCAAGTTCCTGCGGACAACTGCCCCACCTAC 1581
Db 476 -----AlaAlaArgLeuPro-----ProGluArgGlnGluProArgLeuP 489
QY 1582 CTGC-----CGGAGGCGGCTTCGCGCACCTCAAGAGTGGCGGTGGGA 1625
Db 489 roGlnAspLeuAlaAlaGlnArgCysProAlaGlyPro----- 502
QY 1626 CAACCTGACCGTCTCCACCACTTCTTCA-----TGGAGGAGCTCAAGGCGG 1673
Db 503 -----ProProThrArgSerGlyAlaAlaAlaGlnArgThrHisArgArgP 518
QY 1674 CACCTACACCAAGGTGATCTGTGAGCACCAGTGGACTGTGATATGCGCGTGGCCAA 1733
Db 518 roProGly-----CysProArgSerA 525
QY 1734 CGAGCTGGCGAGTGTCTGGCCAGCAGGTGTGGCGGGCGGATCTGTCTGTGGCGCTC 1793
Db 525 laArgAsnProGlyCysProArgThrTrpArgArgArgSerGlyAlaGlnArgGlyHisP 545
QY 1794 CGCTCCCTCAGCGCGGCTTACGCGGAGCTGATCCAGAGCGCGCTTCGACGTGCG--- 1850
Db 545 roProProGlyAlaGlyGlnArgProSer-GlyProThrGlyGlyArgProAlaAlaPro 564
QY 1850 ----- 1850
Db 565 GlyAlaProGlyThrProAlaAlaProGlyProGlyGlyAlaAlaValProSerGly 584
QY 1851 -----CTGCATCGCGCGGCTCTAGGGCTCATGGACCGGCTCAACATGTACAGTCTC 1904
Db 585 AlaThrProHisProGluArgGlySerGlyProAlaAspProProAlaAlaAlaArgLeu 604
QY 1905 CTCTACATGCGCGCGGAGGGCGCCAGAGAGGCAACTA 1946
Db 605 -----ProProGlu---ArgGlnGluProArgLeu 613

RESULT 15
T45462
membrane glycoprotein [imported] - equine herpesvirus 1
C;Species: equine herpesvirus 1
C;Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 09-Jul-2004
C;Accession: T45462
R;Kirisawa, R.; Kobayashi, T.; Kawakami, Y.; Iwai, H.
J. Equine Sci. 7, 79-87, 1996
A;Title: Nucleotide sequences of open reading frames 1, 24 and 71 of an attenuated equine herpesvirus 1
A;Reference number: Z22973
A;Accession: T45462
A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA
A;Residues: 1-866 <IR>
A;Cross-references: UNIPROT:O39781; UNIPARC:UPI00000ECBA1; EMBL:D88733; PIDN:BAA20037.1
A;Experimental source: strain HH1
C;Genetics:
A;Note: ORF71
C;Superfamily: equine herpesvirus glycoprotein X; equine herpesvirus 1 glycoprotein homolog
F;558-866/Domain: equine herpesvirus 1 glycoprotein homology <HG>

Alignment Scores:
Pred. No.: 2,18e-06 Length: 866
Score: 245.50 Matches: 119
Percent Similarity: 39.3% Conservative: 76
Best Local Similarity: 24.0% Mismatches: 204
Query Match: 6.8% Indels: 97
DB: 2 Gaps: 13

US-10-620-914-44 (1-1947) x T45462 (1-866)

QY 473 ACAACGTATCGACACGAGCTTGTCTGCTACCTGCTCCCAAGACGCGCTGTGGCGTTGCCG 532
Db 46 ThrThrAsnSerSerSerProThrThrSerProThrThrSerSerSerProPro 65
QY 533 ACTTCTAGTGAAGCGGCAAGTACGACCTGCGCCCTGGCGCAGATGCCCTGTGGCGCGCTT 592
Db 66 ThrSerThrHisThrSerSerProSerSerThrSerThrGlnSerSerSerThrAlaAla 85
QY 593 TCTTCTGGCGATCGATCTTCGACATCGACACATTCATCGACATCGCGCCGCGCGCGCT 652
Db 86 ThrSerSerSerAlaProSerThrAlaSerSerThrThrSerThrSerThrSerThrSer 105
QY 653 ACCTGAGCAGAGCTGAGCGGTGTGGGAGCAGAACACCCAGAGGTTCGATCCCTACG 712
Db 106 ThrGluThrThrThrThrProThrAlaSerThrThr-----ThrProThr 121
QY 713 TGGCGTGGTGGCGCGCCCTACTAGTGTGGATGTGGCGCCCTGCGCCAGGTGGCC--- 769
Db 122 ThrThrThrAlaAlaProThrThrAlaAlaThrThrAlaValThrThrAlaAlaSer 141
QY 770 -----ACGCGCTGACGAGGAGCGGTGGCGCGC 802
Db 142 ThrAlaAlaSerThrSerAlaGluThrThrThrAlaThrAlaThrAlaThrSerPro 161
QY 803 CCATGTTTCCGCGCCCTCTCTGTACACGACGCTGTGGAGGAGACCCGCGCGGATATGG 862
Db 162 ThrThrThrThrProThrSerThrThrThrThrAlaThrThr----- 176
QY 863 AGGTGATGAGATCAACCCCAAGGACACGGTGTGACCCCTGACTAGCGCGCGCTGCAATG 922
Db 177 -----ThrValProThrThr----- 181
QY 923 CCTGAACCTGTGTGGTGCAGGCGCGCGCGGTGTGTGGTGGACTGCAACCCCGCGC 982
Db 181 ----- 181
QY 983 AGTGGCGCTTCTGGAGCTGAAGAGGTGGCCATTCAGCAGCTGGAGTTTGGAGACGTGT 1042
Db 182 -----AlaSerThrThrThrAspThrThr 189
QY 1043 GGCAGCTGTGGCGAGGCGGTGCACCCGCGCATGAGGAGCTGTACGAGAGAGCTGG 1102
Db 190 ThrAlaAlaThrThrThrAlaAlaThrThrThrAlaAlaThrThrAlaAlaThrThr 209
QY 1103 CGCCCTTCTGTGCGCAACACCGCCCACTTCTGGTCCAGCGCCCTCTGTACTTCCAGC 1162
Db 210 ThrAlaAlaThrThrThrAlaAlaThrThrThrAlaAlaThrThrThrAlaAlaThrThr 229
QY 1163 ACGCGCTGTACTACAGGCGCGCATGGCAAGCTGTGGTGGTGGTGGTGGTGGTGGTGGC 1222
Db 230 ThrAlaAlaThrThrThrAlaAlaThrThrThrAlaAlaThrThrThrSerSerAlaThr 249
QY 1223 TGGTGTGGGACTGGGCAAGACCGTCAAGCGCTCGCCCAACCGCCCAACCAATGGAGAGC 1282

```
Db      250 AlaAlaThrThrAlaAlaThrThrAlaAlaThrThr----- 263
Qy      1283 AGGCCGCTCTGGGACAGCAACATGCTCATCCACTTCGTGAAGACGGGCCCAAGCCGC 1342
Db      264 -----ThrAlaAlaThrThrAlaAlaThrThr-----ThrAlaAlaThrThr 278
Qy      1343 TGGTGTGGCTGTTGGTCAAGTTCGTGAGCCCTGCTGCTCTTCAACAAGGCGGTGCTGGT 1402
Db      279 ThrAlaAlaThrThrAlaAlaThrThrAlaAlaThrThrAlaAlaThrThr 298
Qy      1403 TCGCGCGCGCGTCCGGGCAAGCAGTACCGCTGATCAAGCGGAGCGGATCCCCATTG 1462
Db      299 ThrAlaAlaThrThrAlaAlaThrThrAlaAlaThrThrAlaAla----- 316
Qy      1463 AGAATACATCGCGCGCACCATGAGCGGCTGGGGAGAACTCGCAGCTGCGCAAGCAGA 1522
Db      317 ---ThrThrAlaAlaThrThrAlaAlaThrThrThrGlySerProThrSerGly 335
Qy      1523 ACTACTTCTACTACA-----ACTGCCTCACCGGCAAGTTCCTGCGGCACTGCC 1573
Db      336 SerThrSerThrThrGlyAlaSerThrSerThrProSerAlaSerThrAlaThrSerAla 355
Qy      1574 CCACCTACCTCGCGGAGCGGCTTCGCCACCTCAAGAGTGGCGTGGTGGAACAACCTGA 1633
Db      356 ThrProThr-----SerThrSerThrSerAlaAlaAlaThrThrSerThr 370
Qy      1634 CGGTCTCCCAACT-----TCTTCATGGAGAGCTCAAGCGCGCACCTACACCAAGG 1687
Db      371 ProThrProThrSerAlaAlaThrSerAlaGluSerThrThrGluAlaProThrSer--- 389
Qy      1688 TGATTCTGATGGACCACTGGAGCTGGCTGGATATGCCCGTGGCCACGAGCTGGCGGAGT 1747
Db      390 -----ThrProThrThrAspThrThrThrProSerGluAlaThrThrAlaThr 405
Qy      1748 GCCTGGCCAAGCAGGTTGCGCGGCGGCGCATCGTCACTCGCGCTCCGCTCCCTCAGCC 1807
Db      406 ThrSerProGluSerThrThrValSerAlaSerThrThrSerAla---ThrThrThrAla 424
Qy      1808 CGCCTACGCGGAGCTGATCCAGAGGCGGCGTTCGACGTCGCGCTGCATCGCGCGGCCA 1867
Db      425 PheThrThrGluSerHisThrSerProAspSerSerThrGlySerThrSerThrAlaGlu 444
Qy      1868 CTCAGGCTACATGACCGCGGTCAACATGTACAGCT---CCTTCTACA 1912
Db      445 ProSerSerThrPheThrLeuThrProSerThrAlaThrProSerThr 460
```

Search completed: March 14, 2006, 01:45:57
Job time : 82 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 14, 2006, 01:44:41 ; Search time 7.6 Seconds
(without alignments)
4236.041 Million cell updates/sec

Title: US-10-620-914-44

Perfect score: 3619

Sequence: 1 atggggctcggttgctgacgg.....gcccgaagaagacaactaa 1947

Scoring table:

BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 1144120

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlh

-O=/abs/ABSSWEB.spool/US10620914/runat_13032006_101932_5572/app_query.fasta_1

-DB=Issued Patents AA_QFMT=faetan -SUFFIX=n2p_rai -MINMATCH=0.1 -LOOPEL=0

-LOOPEXT=0 -UNITS=BITS -START=1 -END=1 -MATRIX=bloms62 -TRANS=human40.cdi

-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15

-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000

-HOST=abs05h -USER=US10620914 @CGN 1 1 33 @runat 13032006 101932 5572 -NCPU=6

-ICPU=3 -NO MMAP -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG -DEV TIMEOUT=120

-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7

-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents AA:*

1: /cgn2_6/ptodata/1/iaa/5 COMB.pep.*

2: /cgn2_6/ptodata/1/iaa/6 COMB.pep.*

3: /cgn2_6/ptodata/1/iaa/H COMB.pep.*

4: /cgn2_6/ptodata/1/iaa/PCTUS COMB.pep.*

5: /cgn2_6/ptodata/1/iaa/RE COMB.pep.*

6: /cgn2_6/ptodata/1/iaa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result

No. Score Query Match Length DB ID Description

c 1 374.5 10.3 663 2 US-09-252-991A-30843

c 2 352 9.7 310 2 US-09-248-796A-21233

c 3 332.5 9.2 663 2 US-09-252-991A-30843

c 4 332 9.2 536 2 US-09-252-991A-16754

c 5 326.5 9.0 681 2 US-09-252-991A-24567

c 6 324 9.0 2294 2 US-09-252-991A-17231

c 7 323 8.9 638 2 US-09-252-991A-27068

c 8 317.5 8.7 467 2 US-09-252-991A-18296

c 9 316.5 8.7 1706 2 US-09-252-991A-31760

c 10 310 8.6 1073 2 US-09-252-991A-27341

c 11 309 8.5 1064 2 US-09-252-991A-23774

c 12 308 8.5 1064 2 US-09-252-991A-17508

ALIGNMENTS

RESULT 1

US-09-252-991A-30843

; Sequence 30843, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; PRIOR FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 30843

; LENGTH: 663

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-30843

Alignment Scores:

Pred. No.: 2.18e-19

Score: 374.50

Percent Similarity: 33.2%

Best Local Similarity: 28.6%

Query Match: 263

DB: 49

US-10-620-914-44 (1-1947) x US-09-252-991A-30843 (1-663)

QY 1771 CCGGCGCAACCTGCTGGCCAGGCACTCGGCGAGCTGTTGGCCA-----CGGCATAT 1718

Db 4 ProAlaProAlaP-----ProGlnArgTrpProArgGAlaAa 19

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```
QY 1717 CCAGCCAGTCCA-----CGTGGTCCATCAGATCA----- 1688
Db 20 ArgAlaSerProTrpSerThrThrArgProProAlaAspArgSerGlyAlaThrAla 39
QY 1687 -----CCTTGGTGT-----AGTGGCGGCTTTGAGTCTCT-----CCA 1655
Db 40 ArgValProAlaCysArgArgValProIleArgCysAlaAsnAlaTrpProGlyArgPro 59
QY 1654 TGAAGAAGTTGGTGAGACGGTCAGGTGTTCACACGCCACCTCTTGAGGG----- 1604
Db 60 MetSerSerThrSerArgArgProGlyTrpSerProAlaValProAlaGlyAlaCysCys 79
QY 1603 TGGCGAAGCGCGCTCGCAGGTAGG-----TGGGCGAGTTGTGCGCA 1559
Db 80 TrpLysThrArgAsnValAlaGlyArgSerAlaThrAspAlaTrpSerSerAlaProAla 99
QY 1558 ---GGAACCTTTCGGGTGAGCGAGTTGT-----AGTAGA 1529
Db 100 ProAlaAsnCys-----CysCysLeuPheProAlaGlyProSerProAlaSerPro 116
QY 1528 AGTAGTTCTGCTTGGCAGCT-----GCGAGTTCTCGGCCACGCCCTCCATGG 1481
Db 117 AlaProAlaAlaCysArgArgTrpProArgAlaAlaCysHisTrpProAlaSerAlaTrp 136
QY 1480 TGC-----GCGCGATGTAGTTCTCAATGGGGATGCCGTCG---CCTTGATCA 1436
Db 137 TrpTrpLeuAlaProAlaArgCysCys-----TrpProAlaProAlaProAlaSer 154
QY 1435 GCGGTACTGCTTGGC-----CGGCAACGCGCGCGCGAACCACGACGCGCTTGT 1385
Db 155 AlaGlyArgAlaCysCysAlaSerProAsnArgArgArgGluProTrpProPro--- 173
QY 1384 TGAAGACACAGGCTCAGCACTTGACGAACGCCACACCGAGGCTTGGGCGGCTTCT 1325
Db 174 -----SerProTyAla----- 177
QY 1324 TCAGAGTGCATGAGCATGTGTGTGCCACAGAGCGGCTGTCTCTCCATTGTGGCG 1265
Db 178 SerArgAlaGlyProAlaSerCys-----GlyArgProPro----- 189
QY 1264 CGTTGGCGAGCGCTTGACGGTCTTGGC-----CCAGTCCCAGCACCGCCGAGGCACTGCA 1208
Db 190 -----AlaCysSerProValAlaThrAlaProAlaThrAlaThrCys 202
QY 1207 GCACCCAGCACAGCT-----TGCCCATGCCGCCCTGTGTAGTACAGGC 1166
Db 203 SerProProSerAlaArgSerAlaTrpLysProCysAlaCysAlaLys---AlaValGly 221
QY 1165 CGTGTGGAAGTACCAGAGGCGCTTGACAGGAAGTTGTGGCTGTTGGCAGAGGAAG 1106
Db 222 SerAlaArgSerProAlaAsnAlaTrp----- 230
QY 1105 GCGCCAGCTTCTTCTGTACAGTCTCTCAATGCGCGGGTGACGCGCTCGCCGCAACAGCT 1046
Db 231 ---ProAlaAlaSerAlaTrpCysProThrCysAsnSerAlaArgProSerAlaIleAla 249
QY 1045 GCACACA-----CGTCTCT-----CAAACT 1028
Db 250 SerThrAlaArgArgSerProSerThrAsnGlyArgProAlaAlaCysProThrThrThr 269
QY 1027 CCAGCTGCTGAATGGCCACCTTCTTACGCTCCAGAAGCGCGACTGCGCGGGTTGCAGT 968
Db 270 ProProAlaAsnAlaProAlaSerAlaAlaAlaAsnTrpProTrpTrpLysAlaProSer 289
QY 967 CCACCGACACACCTGGCGCGCCCTCGCACCA---GCAGGTTCCAGGCGCATTCGACCGCG 911
Db 290 ProAlaMetProProSerThrSerAlaThrArgLeuAlaAlaCysGlyHis----- 306
QY 910 CGTAGTCAGGTCAGCA-----CCGTGT 887
Db 307 ---AspGlyAspAlaGlyArgAlaSerGlnThrProTrpArgAspThrSerProCys 324
QY 886 -----CCTTGGGGT-----TGATCT 872
```

```
Db 325 ValProAsnCysValAsnTrpArgArgArgGilleProTrpTrpSerAlaValAlaLysThrSer 344
QY 871 CMATCACTCCATATCCGGCT-----CGGGTCTCT---CCCACGACTGGGTACAGGA 821
Db 345 ProTrpProArgTrpProGlyMetProAlaGlyProArgProSerCysThrAlaValAla 364
QY 820 AGGTGGGCG-----GGAACATGGCGCGCTCCACGC---GCTCTCTGTCGAGGG 773
Db 365 AlaTrpAlaProAlaArgGlyGlyTyAlaAlaAlaProArgAsnSerCysSerAlaGly 384
QY 772 CGTGCCCAACGC-----TGGCGAGCGCGC 749
Db 385 ArgLeuArgArgArgValArgArgSerAlaArgAlaTrpArgProTrpProAlaGly 404
QY 748 CAATCCACA-----CGTAGTAGGGGGCGCGCGACCG 716
Db 405 ArgAlaThrProAlaSerAspAlaArgAspSerAlaSerArgGlyArgProAlaThr 424
QY 715 GCAGGTAGGGGATCGAACCCTGGGTGTCTGTCTCCACACGCGCTCCAGCTTCTGTCTCCA 656
Db 425 AlaAlaAspHisProAlaAlaTrpValSerAla---AlaArgArgThrSerSerAlaPro 443
QY 655 GSTAGGCGCGCGCTCGGGCCGATGTCATGTGTGTGATGTGCAAGATCGATCGCCAGA 596
Db 444 IleAlaGlySerAlaProGlySerGlyThrAlaProArgCysHis----- 458
QY 595 AGAAACGGCGCGACAGGCGATCTGGCGCAGGGCAGGTCTGCTACTTCCGCTCAGCTAGA 536
Db 459 -----ProValArgLysAspGlyAlaGlyProAlaSerThrGlyArgSerArgArg 475
QY 535 AGTCGCGCAACGCCACCGCGCTTGGGACAGGTACGAGCAAGCGCTGCTCGATGACGT 476
Db 476 ArgTrpAlaAlaProAlaArg-----ArgAlaArg 486
QY 475 TGTGAACCGTGGAAATCATCTGAGCGAGTGGAGAGGTGATGAGCGTCCGCTGCGCT 416
Db 487 ThrGlyIle-----ArgArgThrAlaAlaAlaCysTyr 498
QY 415 CAGGGGGCGCAATTTGCAAGCGTCCGCTCCA----- 383
Db 499 ProProProAlaProGlyArgArgThrProAlaAlaArgArgArgSerAlaArgThrAla 518
QY 382 -----CGACCTGGACATTTCTCCAGCGCTTCCGCTTCCGCTTCCGCT 347
Db 519 ProArgProAlaIleProAlaHisArgProSerAlaSerAlaThrGlyTrpProGlyPro 538
QY 346 TCTTCTTGGCCACTCGCACGAGGTGGCACAGGTGCGACCGTAGATGGACTTGA 287
Db 539 AlaLeuAlaCysProAlaAlaAla----- 546
QY 286 TCGCCAGGTGATGTAATCAGCCATCATATCGACATTTCTCCCGAGTGCCAC---CACCCA 230
Db 547 -----GlyArgArgProSerProAlaProAlaProAlaProAlaProAlaArgPro 564
QY 229 GGTCAACCCAGATGAGGTTTCGAGCGCTCGGCCGCGCGGCGAGCAAGCGCGCTGGG 170
Db 565 ArgAlaProArg-----ArgArgProGlyArgSerProAlaThrAlaProAla 580
QY 169 GCCCGTAGAAGCTTCTCCAGCGAGCAGGTGATCATCGCTTCTTCTGTCGCCGAACCA 110
Db 581 Ala---PheAlaAlaProAlaGlyProThrProAlaArgProProAlaAlaAlaThr--- 598
QY 109 TATGGCGCAGAACGGTCAGGTCTCTTTCATGTGTGCTGCTGAGCTTCTCCAGGGAGA 50
Db 599 -----GlyArgProSer-----AlaProProArgProAlaArg 609
QY 49 AGTTCT-----TCTTGTGTAGTCTCGAGCGCGCGCTCACGACCCGACCCCA 2
Db 610 ArgArgValArgTrpProTrpArgSerAlaArgCysSerProArgArgProThrPro 628
QY 628
```

RESULT 2

US-09-248-796A-21233

QY	868	ATGAGATCAACCCCAAGGACACGGTGTGACCCCTGACTAGCGCGGCTGCAATGCCCTG	927
Db	215	LeuAsnPheThrSerAspThrValLeuAlaIleThrSerAlaGlyAspAsnIleLeu	234
QY	928	AACCTG-----CTGGTGACGGGGCGGCGGCGTGTTCGGTGGACTGCAACCCCGCG	981
Db	235	SerTyrAlaSerLeuProThrProProlYsIleHisAlaValAspLeuAsnProCys	254
QY	982	CAGTCGGCGCTCTCGAGCTGAGAAAGGTGGCCATTTCAGCAGCTGGAGTTTGAGGACGTG	1041
Db	255	GlnAsnHisLeuLeuGluLeuLysLeuAlaSerPheArgCysLeuSerGlnGluGlnIle	274
QY	1042	TGCGACCTGTTTCGGCAGGGCGGTGCACCCGCGCATTTGAGGAGCTGTACGAGAAGAAGCTG	1101
Db	275	TrpSerMetPheGlyGluGlyIleGluAsnPheAsnAspLeuLeuIleAspThrLeu	294
QY	1102	GCGCCCTTCTCTGCGCAACACCCAGCCACCACTTCTGGTCCAAGCGCTCTGGTACTTCCAG	1161
Db	295	AlaProHisMetSerSerAsnAla-----PheGln	304
QY	1162	CAC 1164	
Db	305	His 305	

RESULT 3
 US-09-252-991A-30843
 ; Sequence 30843, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 30843
 ; LENGTH: 663
 ; TYPE: PRT
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-30843

Alignment Scores:		
Pred. No.:	2,81e-16	Length: 663
Score:	332.50	Matches: 193
Percent Similarity:	32.4%	Conservative: 52
Best Local Similarity:	25.5%	Mismatches: 253
Query Match:	9.2%	Indels: 259
DB:	2	Gaps: 44

US-10-620-914-44 (1-1947) x US-09-252-991A-30843 (1-663)

QY	29	CGAGCTACACCAAGA-----AGAACTTCTCCTGAGAGACTCA-----	67
Db	12	ArgTrrProProArgArgAlaAlaArgAlaSerProTrrSerThrThrArgProPro	31
QY	68	-----AGCTCAGCAGCATGAAGGAGACTGACCGTTCGCGCCATATGTGTTG	118
Db	32	AlaAspArgSerGlyAlaThrAlaArgVal-----ProAlaCys-----	44
QY	119	GCAGCAAGAAGCGCGATGATCAGCTGCTC-----GCTGG-----AGAGTTCT	163
Db	45	-----ArgArgValProIleArgCysAlaAsnAlaTrrProGlyArgProMetSer	62
QY	164	ACGGGCCCCAGCGCTGCTTTGTGTCGGCGCTGGCCAGCGCTCGAACCTCATCTGG	223
Db	63	ThrSerArgArg-----ProGlyTrrSerProAlaValProAlaGlyAla	77

[illegible]

```
QY 358 TGGCTTGGCTTCTTCTTGGCCACCTCGCACAGCGAGTGGCACAGGTGCACACCGACGTAGA 299
Db 499 TTPProProProGlySerProThrGlyAlaArgArgThrThrSerArgCysArg 518
QY 298 TGGACTTGAACCTTCGCCAGGT 278
Db 519 TTPCysGlyArgThrProGly 525

RESULT 5
US-09-252-991A-24567
; Sequence 24567, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24567
; LENGTH: 681
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24567

Alignment Scores:
Pred. No.: 7.88e-16 Length: 681
Score: 326.50 Matches: 214
Percent Similarity: 32.8% Conservative: 38
Best Local Similarity: 27.8% Mismatches: 262
Query Match: 9.0% Indels: 255
DB: 2 Gaps: 50

US-10-620-914-44 (1-1947) x US-09-252-991A-24567 (1-681)
QY 1873 CCTGAGTGGCGCGCGGATGC---AGCGCAGTTCGAAGCCGCT-----TCTGGA 1826
Db 10 ProProTrp-----AlaCysAsnSerThrGlyArgThrProThrAspGlyArgSerGly 27
QY 1825 TCAGCTGGCGTAGGGCGCGCTGAGG-----AGCGGAGCGCCAGA 1784
Db 28 SerThrArgArgSerSerAlaAlaCysTrpLysHisSerAlaIleArgArg-gValArg 47
QY 1783 TGACGATGCCCGCGCGGCAACCTGCTTGGCCAGC-----ACTCGGCCA----- 1739
Db 48 AsnArgSerAsnArgAlaTrpProArgTrpProGlyCysAlaArgThrAlaProThrSer 67
QY 1738 -----GCTCGTTGG---CCACGGGCAATCCAGCCAGTCCA 1706
Db 68 AspCysTrpSerAlaAsnAlaAlaSerArgTrpSerLysArgSerAlaProAlaPro 87
QY 1705 CGTGCTCATCAGNATCACCTTGTGTAGG-----TGCGGCTTTGAGCTCTCTCCA 1655
Db 88 ArgAlaSerTrpSerThrArgThrAlaArgGlySerProCysAla----- 102
QY 1654 TGAAGAAGTTGGTGACAGCGTCAAGTTGTCCACCGCCACTCTTGAGGCGTGCGCAAGG 1595
Db 103 -----TrpMetProLysAlaAlaCysArg-----ArgArg 112
QY 1594 CCGCTCCGCGAGGTAGTGGGCGAGTTGTGCGCGCAGAACTTCCGGTGAGGCGAGTTGT 1535
Db 113 ProAlaAlaAlaMetArg-----SerCysArgTrpSerSerGly 125
QY 1534 AGTAGAGTAGTTCTGCTGGCAGCTCGGAGTTCTCGG---CCACCGCTCCATGTGTC 1478
Db 126 AsnGlyArgTrpArgTrpArgArgAlaValArgProTrpProAlaSerProArgVal 145
QY 1477 GCGCGATGTAGTTCTCAATGGGGATGCGCTCGGCTTGATCAGCGCGTACTGCTTGCCCG 1418
```

```
Db 146 AlaArgValAlaGlyValTrp-----ProPro-----SerLeuCysPro 158
QY 1417 GCACGCGCGCGCGGACCAACACAGCAGCGCTTGTGAAGAGACACAGGCTCAGCAACTGA 1358
Db 159 AlaProProArgArg-----ArgArgProGly---ArgGlnArg 170
QY 1357 CGAACAGCCACACAGCGGCTTGGGCGCTTCTTTCACGAAGTGGATGAGCATGTTGCTGT 1298
Db 171 ArgProArgGlyProAlaAlaGlnArgArg----- 180
QY 1297 CCCACAGACGCGCTGCTCTCCATTTGGGCGGCTTGGCGAGCGCTTCAGCGCTTTCG 1238
Db 181 -----AlaAlaTrpArgArg-----ArgAlaGly 188
QY 1237 CCAGTCCCAGCACCAGCGCCAGC-----ACT 1211
Db 189 AspGlnProAlaAlaArgProGlyArgGlyGlnArgThrCysLeuGlnProLeuLeuAla 208
QY 1210 GCAGCACCAGCACACACACACACACACACACACACACACACACACACACACACACAC 1181
Db 209 ValGlnProAlaValLeuGlnCysProAlaAlaArgAlaGlyAspHisSerArgArgGly 228
QY 1180 -----CCTGGTAGTACAGGCCGT 1163
Db 229 SerGlyGlyThrGlyAspProSerArgArgProGlyArgGlyAspGlyProProGlyGlu 248
QY 1162 GCTGGAAGTACACAGCGCTTGGACACAGAGTTGTGGCTGTTTCCGACAGCAAGGGCG 1103
Db 249 ProGlyAlaAspArg-----LeuAspSerGlyGlyGly 259
QY 1102 CCAGCTTCTTCTGCTACAGCTCT-----CAATGCGCG 1070
Db 260 ProAlaHisAlaProAlaAlaProThrProArgLeuHisArgAlaArgGlnSerLeu 279
QY 1069 GGTGCAAGCGCTCGCGAACAGCTGCCACACGCTCTCAACTCCAGCTGCTGAATGGCCA 1010
Db 280 AlaProArgProArgArgLysProProArgSerArgArgGlyProAlaAlaProLeuPro 299
QY 1009 -----CCTTCTTCAGCTCCAGACGCGCGCTGCGGGTTCG 971
Db 300 LeuArgAspProAlaGlyAlaProGlyArgArgProArgLeuThrLeuAlaGlyAla 319
QY 970 AGTCCACCGACACACACCTGGCGCGCGCTGACACAGCAGGTTTCAGGCGCAATGC----- 917
Db 320 Ala-----AlaProProArgArgAlaGlyGlyGlyValLeuArg 333
QY 916 ---AGCGCGCTAGTTCAGGTCAGCACCGCTGT-----CCTTGGGGTTGATCTCCATCA 866
Db 334 ArgSerArgArgGlySerGly---LeuProCysLeuArgPro-----ValAlaAsp 349
QY 865 CTTCATATCCGCTCGGGTCTCTCCACGACCTGCG-----TGTACA 824
Db 350 ProThrLeuProAlaThrCysProAlaProGlyAlaArgGlyArgAspGlyHisArgSer 369
QY 823 GGAAGTGGCGCGGA-----ACATGGCGCGCGCTCCA 791
Db 370 GlyArgArgProGlyGlyArgArgArgGlnProGlyLeuValAlaAlaGly 389
QY 790 CGCGCTCTGTCGAGGCGCTGCCAACCGTGGGCGGCGCG----- 749
Db 390 ArgAlaAlaArgSerGluCysArgArgAlaAlaGlyHisProGlnProValArgThr 409
QY 748 -----CAATCCACACCT----- 737
Db 410 GlyLeuGlyArgPheArgValGlnSerArgArgProAlaProArgLeuProGlyIle 429
QY 736 -----AGTAGGGGCGCGCACCGCCAGCAGTGGGATCGAACCCCTGGGTGTCT 686
Db 430 ProArgAspAlaAlaGlyGlnProGlyLeuAlaArgAlaThrHisArgProCysAsp 449
QY 685 GCTCCCAACACGCGCTCAGCTTCTGCTCCAGTAGGCGCGCGCT---CGGGGCGGATGT 629
```

```

Db 450 GlyProAlaThrSerLeuAlaAsp---ProAlaArgProAlaAlaProAlaArgArgLeu 468
Qy 628 CAATGTTGTCGATGTCGAAGATCATGCCAGAGAAACGCGACCGACCGATCTGGC 569
Db 469 ProAlaLeuProAlaThrArgAlaAlaProAla---GlyAlaGlyGlnSerArg 487
Qy 568 GCAGGGCAGGTCGATCTGCGCTCAGCGTAGAAGTCGGCAACGCCACCGAGCGCTTT 509
Db 488 LeuGlyAlaGlyHisArg---ArgGlyProArgHisArgAlaArgArgProAlaArg--- 505
Qy 508 GGGACAGTCAGACAGCCTGCTCGATGATGCTGTGTGGAACGGTGGATCATCGTAGCG 449
Db 506 GlyThrGlyThrProPro-----SerAlaGlyTyrProArgAlaAlaLeuArg 521
Qy 448 AGTAGGAGAAAGGTGATGAGCGTGCCTGCGCTCAGGGGGCGCAAAATTTGGCAAGCGTCGG 389
Db 522 AlaAlaArgArg-----ThrLeuArgAlaPro---GlyAlaMetAlaGlyArgHis 538
Qy 388 CTTCCAGCAGCTGACATCTTCCAGCCCTTGGCTTGGCTTCTTCTTGGCCACCTCGC 329
Db 539 -----GlyHisHisGlnHisProArgProAlaGlu-----ProValArg 551
Qy 328 ACAGCGAGTCGACAGCT-----CGACCAGTAGATGAGCTTGAATTCGCCCA 281
Db 552 LeuValAlaGlyProGlyHisProLeuAlaArgProArg-----ThrPro 567
Qy 280 GGTGCGATTAATCAGCCATCATATCGACATCTCCCGCAGTCGCCACCCAGGTCAACCC 221
Db 568 GlnArg-----ArgGlyMetArgGlyArgProGlyThrAlaArgArgGlyAlaPro 584
Qy 220 AGATGAGTTGAGCGCTCGCGAGCGCGGAGCAAGGAGCGGCTGGGGCCCGTAGA 161
Db 585 GlyProGlyArgGlnProArgThrAlaGlyArgSerArgProArg-----ArgArg 601
Qy 160 AGCTCTCCAGCGCAGCGATGATCGC-----CGTTCTTCTGCGCGCA 116
Db 602 SerThrAlaGly-----ArgLeuHisArgLeuCysArgGlyAsnProArgAlaValGly 619
Qy 115 ACCACATATGCGCAGAACCGTCAGGTTCATCTTCATGCTGC-----74
Db 620 AlaThrAlaProGlyArgProGlyGlnProGlyAlaAlaGluProAlaArgProArg 639
Qy 73 -----TGAGTTGAGCTTCTCCAGGAGAAAGTTCTTCTTGGTGTAGCTCG 29
Db 640 ArgArgProProGluLeuAlaAlaLeuAlaGlyLysArgArgAlaAspAlaArgAsn 659
Qy 28 CAGCGCGCGCTCAGACCCGCCCA 2
Db 660 AlaAlaGlyArgProAlaProAlaPro 668

```

RESULT 6

```

US-09-252-991A-17231
; Sequence 17231, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 1998-07-27
; SEQ ID NO 17231
; LENGTH: 2294
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17231
Alignment Scores:

```

```

Pred. No.: 1-97e-15 Length: 2294
Score: 324.00 Matches: 274
Percent Similarity: 29.3% Conservative: 43
Best Local Similarity: 25.3% Mismatches: 280
Query Match: 9.0% Indels: 485
DB: 2 Gaps: 58

US-10-620-914-44 (1-1947) x US-09-252-991A-17231 (1-2294)

Qy 20 GCGGGCTCGAGCTACACCAAGAACTTCTCCCTGG-----58
Db 55 AlaGlyHisArgGlnArgProArgArgAlaAlaGlyTyrTyrAlaValGlnHisProArg 74
Qy 59 ---AGAAGCTCAAGCTCAGCAGCA-----TGAAGGATG 88
Db 75 ArgArgSerGlyAlaGlyAlaAlaGluCysArgGlyLysProAlaArgTyrHisArgLeu 94
Qy 89 ACCTGACCGTTTCGCCCATATGTGTTCCGGCAGCAAGAGGGCGATGATCAGCTGCTC 148
Db 95 ProArgPro---AlaAlaGluAlaGlyArgHisProArgHisGly-----108
Qy 149 GCCTGGAGAGCTTCTACGGGCCCCAGGCCGCTGCTTGTGCGCCGCTGCG-----200
Db 109 -----ArgGlyArgGlyProArgArgProLeuArg-GlnProGlyAlaGlySe 126
Qy 201 -----CGAGCGCTCGAACTCATCTGGGTTCACCTGGTGGTGGTGGTGGTGGT 244
Db 126 rAlaArgArgGlnCysArgLeuArgGluProArgProGlyGlyProGly-----142
Qy 245 GGAGAAATGTCGATATGATGGCTGATTATCATCGACTGCGAAGTTCAGTCCATCTACG 304
Db 143 -----GlnProGlyValProProAlaAspArgAl 152
Qy 305 TGCTGACCT-----GTGCCACTGCTGCGAGG 334
Db 152 aArgArgProArgArgIleArgArgGlyArgAlaGlyProArgGluAlaArgG 172
Qy 335 T---GCCAAGAAAGAGCGAAGGC-----CAAGGCTGGAAGATGT-----374
Db 172 yProGlyProGluProGlyArgGlyLeuProAlaAlaGlyAlaArgHisCysArgArgAl 192
Qy 375 -----CCAGGTCT-----383
Db 192 aAlaProGlyValLeuArgGlnGlyAlaAlaGlyProAlaProProAlaArgSerPr 212
Qy 384 -----GGAGGCGGAGCTTGGCAAT 403
Db 212 oArgArgAlaGlyGlyGlyGlyGlnArgArgGlyArgLeuGlyAsnArgArgLeuProAl 232
Qy 404 TTGCGCCCCCTGAGGCGACCGCTCATACCTTCTCTACTC-----GC 451
Db 232 aAlaAlaThrGlyGlnProArgArgGlyProHisTyrSerProValProArgAlaProAl 252
Qy 452 TCACGATGATCCACCGTTCCACACGTCATCGACGAGCTTGTCTGTA-----500
Db 252 aAlaProAlaGlyAlaValProArgThrGlyGlyProGlyArgArgValLeuTyrLeuHi 272
Qy 501 -----CCTGTC-----CCAAGCGGCTGTTGGCG 526
Db 272 sArgGlnProAlaAlaSerArgGlyValProGlyValGlnProArgArgProGly-----290
Qy 527 TTGCGGACTTCTACGTGAGCGGCAAGTACGACTGCGCTGCGCTGCG-----569
Db 291 ----ArgGlnLeuArgProGlyAspAlaArgProAlaProGlyAlaGlyAspGlyAspAr 309
Qy 570 -----CCAGATGCTGTCGCGCGCTTCTTCTTCTGCGCATCGATCT 610
Db 309 gArgGlnGlyAspArgHisProAspArgArgGluGlyLeuArgArgAlaArgAspAl 329
Qy 611 TCACAT-----CGACACATTTGACATCGGCC-----CGAGCGCC 646
Db 329 aGlnArgProGlnProAspArgGlnArgGlnLeuArgProGlyGlyAlaArgArgArgPr 349

```

647 GCAC-----CTACCTGGAGCAGAGCTGGAGCGCGTGTGGAGCAGA 688
| | | | |
349 oArgArgGluProAlaArgAlaLeuProAlaAlaGlyGlnGly---ArgLeuGlyGlyAl 368
| | | | |
689 ACACCCA-----GGTTC----- 701
| | | | |
368 aHisProArgProGlyGlnProAlaThrAlaGlyTyrArgProAlaAlaAspAlaGlyGl 388
| | | | |
702 -----GATCCCTCA-----CG 712
| | | | |
388 yAlaAlaAlaAspProLeuProArgArgLeuHisLeuPheArgThrGlyProArgGlnAr 408
| | | | |
713 TGCGGTG---GCTGCGCGCCCTACTACTGTGTGGATTGGCCCTGCCAGCGTGTGCC 769
| | | | |
408 gAlaLeuGluAlaAlaAspAlaLeuArgArgLeuProHisArgArgProValPr 428
| | | | |
770 ACGCCCTGCAGAGGAGCGGTGA-----GCGGCCGCCCAT----- 806
| | | | |
428 oArgAlaGluProGlyLeuLeuGlyHisSerArgIleSerAsnAspAlaSerHisArgAr 448
| | | | |
807 -----GTTCCCGCCACCTTCCT-----GT 826
| | | | |
448 gSerLeuArgHisProGlyArgArgArgGlnArgLysProAlaArgProHisHisAs 468
| | | | |
827 ACACGAGTCTGGGGAGGCCCGAGCGGATATGGA----- 863
| | | | |
468 pHisAlaAlaSerArgArgProArgProGlyThrGlyArgProProArgArgse 488
| | | | |
864 -----GGTATGAGATCAACCCCAAGGACACGGTGTGACCTGACTAGCG 910
| | | | |
488 rLeuHisAlaAlaAspGly---GlnPro-----GlyAlaAlaGluArgProAr 504
| | | | |
911 GCGGCTCAATGCCCTCAACTGTGTGTCAGGCGCGCCAGGTGTGTGCTGGTACT 970
| | | | |
504 gArgArg-----ProGlnProAlaGlyAlaGlyArgArgPro---ThrAlaGlyAspAl 521
| | | | |
971 GCACCCCGCGCATC-----GGCGCTTCTGGAGCTGAAGAAGGTGGCCATTCCAGCAGC 1024
| | | | |
521 aHisProProAlaGlnHisArgProSerGlyAlaGlyGlnProAlaArgAlaThrAl 541
| | | | |
1025 TGGAGTTTGGAGGAGTGTGGCAGTGTTCGGCGAGGCGGT---GCACCCGCGCATTCAGG 1081
| | | | |
541 a-----ArgLeuProAlaProValArgGlyGluGlyThrProGlyArgArgGl 557
| | | | |
1082 AGCTGTACGAGAGCAAGCTGCGCCCTCTCTGTCGCAAAACAGCA----- 1127
| | | | |
557 yAlaGlyArgSerAspAlaGlyAlaLeuArgAlaLeuHisArgProArgArgSerGlyAl 577
| | | | |
1128 -----CAACTTCTGGTCCAAAGCG---CCTCTGGTACTTCCAGC 1162
| | | | |
577 aGlnHisAlaValGlyGlnHisGlnArgLeuGlyGlnAlaGluProAlaAspHisProAl 597
| | | | |
1163 ACGGCCGTGTACTACAGGCGCGCATGGG----- 1190
| | | | |
597 aGlnArgGlyLeuGlyArgArgGluGlyLeuProValaAlaArgThrLeuProAlaGluPr 617
| | | | |
1191 -----CAAGCTGTGCTGGTGTGTCAGTGCCT----- 1217
| | | | |
617 oAlaProAlaProAlaProAlaGlyAlaAlaValProValHisGlnProArgLeuArgAr 637
| | | | |
1218 -----GCGCGTGTGCTGGGAC 1234
| | | | |
637 gProLeuProGlyHisAlaArgArgProGlnProThrGlyGlyAlaAlaArgAlaHisAr 657
| | | | |
1235 TGGCAAGACCGTCAAGCGCTCGCAACGCGCCACATGGAGGAGGAGCGCGTGTGT 1294
| | | | |
657 gGlyGlyAspProGlnHisProArgArgValArgAlaArgThrValAlaAlaLeuAlaAr 677
| | | | |
1295 GGGACAGCAACATCTCATCCA-----CTTCGTGA 1324
| | | | |
677 gProGlnArgAspAlaArgProProPheAlaIleProAlaAlaValGlyArgTyrArgHi 697
| | | | |
1325 AGAAGCGGCCCAAGCGCGTGTGTGCTGTTCGTCAAGTT----- 1364

697 sArgProGlyLeuAlaAlaGlyLeuAlaValArgProAlaAspGluAlaGlyArgArgAr 717
| | | | |
1365 -----CGTGA 1369
| | | | |
717 gArgAlaGlyValGlnAlaHisProArgHisArgArgAspProGlyValaGlyHisArgPr 737
| | | | |
1370 GCGTGTGTCTTCAACAAGGC-----CGTGTGTGTGTTCGGCGCGC 1411
| | | | |
737 oProGlyGlyAlaAlaAlaGlyAspArgAlaAlaThrProGlyAlaLeuProArgArgAr 757
| | | | |
1412 GCGTGGCGGCGAACA-----GTACCGCTGTATCAAGGCGGACGGCATCC 1456
| | | | |
757 gHisGlnGlyProAlaGlyGlyGlyGlyArgArgArgProLeuGlyGlyAspHisPr 777
| | | | |
1457 CCA----- 1459
| | | | |
777 oArgArgArgAlaValArgGlnArgGlnArgGlnArgAlaArgValProAlaAlaAl 797
| | | | |
1460 -----TT 1461.
| | | | |
797 aAlaAlaHisArgArgArgProAlaGlnGlyGlnGlyProGlyValaGlyHisArgAlaAl 817
| | | | |
1462 GAGAACTACATCGCGCGCACCATGGACGGCTGCGGAGAACTCGCACGTGC----- 1513
| | | | |
817 eAlaThrThrGlyArgSerProArgCysAlaThrArgProThrGlyAsnCysProArgPr 837
| | | | |
1514 -GCAAGCAGAACTTCTTACTCAACTGCCTCACCGGCAAGTTCCTGCGCGCACTGC 1572
| | | | |
837 oAlaProArgArgSerProThrSerSerAlaArgProProAlaThrProAlaAlaSerPr 857
| | | | |
1573 CCACCTTACCTGCGCGAGG-----CGGCTTCCGCCACCTCAAGAGTGGCG-- 1618
| | | | |
857 oProLysGlyAlaAlaThrProSerArgTrpArgProMetProAlaProLysAlaAlaHi 877
| | | | |
1619 -----TGTGGACAACCT-----GACGCTCCACCC 1644
| | | | |
877 sAlaThrAlaGlyTrp---LysSerProTyrSerArgArgAlaAlaArgAsnGluArgLeuProG 897
| | | | |
1645 AACTTCTTTCAT-----GGAGGAGCT----- 1664
| | | | |
897 InLeuLeuArgAlaLeuGlyGlyAlaGlyAlaArgProAlaArgProGluProAspHisL 917
| | | | |
1664 ----- 1664
| | | | |
917 euValProArgProAlaAlaAlaProGlyGlnLeuArgThrValGlyLeuGlyAspGlnP 937
| | | | |
1665 -----CAAAGCGCGCACCC 1677
| | | | |
937 roLeuGlyAlaAspArgProAlaValProGlyLeuAspArgLeuProArgAlaAlaHis- 956
| | | | |
1678 TACACCAAGGTGATTCTGTATGGACCCAGCTGGCTGGATATGCCCGTGGCCAAAGCAG 1737
| | | | |
957 --ArgAlaGlyProAlaGlnArgGlyGlyHisAlaGluSerArgArgGlyGlnArgA 976
| | | | |
1738 CTGSCCGA----- 1745
| | | | |
976 laGlyArgGlnArgArgHisArgArgAlaGlyAspProGlnAlaThrHisGlyA 996
| | | | |
1746 -----GTGCTGGCCCAAGCAGGTTCGCGCGCGCGCATCGT----- 1781
| | | | |
996 rgGlyProGlyProAlaGlnAlaGlyGlnAlaGlyArgGlnArgAlaAlaGlnProValA 1016
| | | | |
1782 -----CATCTGGCGCTCCGCTCCCTCAG----- 1805
| | | | |
1016 rgValAlaLeuValArgAspHisArgProAlaArgPheGlyGlnAspHisArgAlaAspG 1036
| | | | |
1806 -----CCGCGCCCTA-----CGCCGAGCTGATCCAGAGGC- 1835
| | | | |
1036 luLeuArgProGlyLeuProAlaArgArgThrAspGlyArgArgGlyAspProArgGlyA 1056
| | | | |
1836 -----GGGCTTCGACGTGCGTGCATCCGCGCGCACCTCAGGGGTACATGAC 1884
| | | | |

QY 475 TGTGAACGGTGGAAATCATCGTGCAGCGAGTAGGAGAAGGTGATGAGCGTCGCGTGCCCT 416
DB 487 -----
QY 415 CAGGGGGCGCAAAATGTCAGAGCGTCGGCCCTCCAGCAGCTGGACATCTTCCAGCCCTTGG 356
DB 488 -----GlyValGlyAspHisArg-
QY 355 CCTTCGCGCTTCTTCTGGCCACCTCGCACAGCGAGTGGCACAGGTTCGACACCGTAGATGG 296
DB 494 -----GlnArgAlaAspGlyAspProLeuAspArg-
QY 295 ACTTGAATCTCGCAGCGTCGATGTAATCAGCCATCATATCGACATCTCCCGAGTGCAC 236
DB 504 -----ArgThrAspGlnProAla-
QY 235 CACCCAGGTCAACCCAGATGAGGTTCGAGCGCTCGGCCAGCGGGGCGAG-CAAAGCGAG 179
DB 515 HisArgGlyArgProArg-ArgArgProGlyProArgLeuArgArgGly 530
QY 178 CGGCTGGGCGCCGTAGAACCTCTCCAGCGCAGCGTGTATCATCGC----- 131
DB 531 ArgArgGlySerProLeuAlaGlyProAlaHisArgArgHisArgGlyAspProAla 550
QY 130 -----CCTTCTTGTGTCGGAACCAACCATATGCGCGCAGAACCGTTCAGGTTCATCCT 83
DB 551 AlaHisArgProProAlaThrGlyArgAlaAlaIleGlyGlyHisAlaGlnProGly 570
QY 82 TCATGCTGCTGAGCTTCCAGGCTTCCAGGGGAGAAGTCTTCTTGGTGTAGTCTGCAGGCC 23
DB 571 ArgAlaCysArgThrHisArgArgThrGlyArgArgGlyGlyArgArgAlaGlyArgSer 590
QY 22 GCGCGTCACGACCCG 8
DB 591 GlyArgArgPro 595

RESULT 8

US-09-252-991A-18296
; Sequence 18296, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 18296
; LENGTH: 467
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18296

Alignment Scores:
Pred. No.: 3,14e-15 Length: 467
Score: 317.50 Matches: 178
Percent Similarity: 32.7% Conservative: 26
Best Local Similarity: 28.5% Mismatches: 198
Query Match: 8.8% Indels: 222
DB: 2 Gaps: 34

US-10-620-914-44 (1-1947) x US-09-252-991A-18296 (1-467)

QY 1828 GGATCAGCTCGCGGTAGGCGGGCTGAGGGAGGGCGGCGCAGATGACGATGCGCGCCG 1769
DB 3 GlyCysAlaGlyCysProAlaGlyCysGlyArgArgGly----- 15
QY 1768 GCGCAACCTGCTTGGCCAGGCACTCGGCCAGCTCGTGTGGCCACGGGCATATCCAGCCAGT 1709

DB 16 -----AsnTrpProAlaArgAlaGlyPheProPro-----ProAlaCys 31
QY 1708 CCACGGTGGTCCATCAGAAATCCTCGTGTAGTGGCGCGCTTGTAGCTCCTCCATGAAGA 1649
DB 32 ArgProGlyThrGlySerSerArgTrpAsnArgAlaGly----- 44
QY 1648 AGTTGGTGGAGACGGTCAGGTTGTCCACCACCGCCTCTTTGAGGGTGGGAGGCGCCCT 1589
DB 45 -----ArgAlaArgGlyCysPro-----GlyCysAlaAlaProPro 56
QY 1588 CGC-----CGAGTAGTGGGCGAGTGTTCGGCGCAGGAATTCCTCCGCTGAGGC 1541
DB 57 AlaGlyValProGlyAlaGlyArg-----SerValProAlaProAlaCysArg- 72
QY 1540 AGTTGTAGTAGAGTAGTCTCTGCTTGGCGACGCGAGTCTTCGCCACCGCTCCATGG 1481
DB 73 -----ArgArgArgHisArgArgArgSerArgProSerSer-----GlyArgProArg 88
QY 1480 TGGCGCGCATGTATCTTCAATGGGATGCGCTCCGCTTCAGTATCAGCGCGTACTGTGTC 1421
DB 89 GlnGlyArgArgGlySer-----ProThrAlaAlaLeuArgLeuArg 102
QY 1420 CCGGCACGCGCGCGCGCAACCCAGCACGCGCTTGTGAAGACGACCGGCTCACGACT 1361
DB 103 ProAlaArgArgArg-ArgHisPro----- 111
QY 1360 TGACGAACACGACCCAGCGGCTTGGCGCGCTTCTTCCAGAGTGGATGAGCATGTGTC 1301
DB 112 -----AlaTrpArgArgArgSerAlaSerArgProAla- 122
QY 1300 TGTCCACACAGCGCGCTCTCTCCATTTGGCGCGCTTGGCGAGCGCTTGACGCTCT 1241
DB 123 -----GlyAlaProProPro- 127
QY 1240 TGGCGAGTCCAGCACCGCGCGCAGGCTGCGACCCAGC----- 1159
DB 128 AlaProCysProAlaProArg-----GlyAlaGlyAlaProSerValArgGluArgProAla 146
QY 1198 ---ACAGCTTGGCCATGCGCGCTGGT-----AGTACAGGC 1166
DB 147 GlySerAlaProProProProAlaArgArgProProAlaProProHisArgThrGly 166
QY 1165 CGTGTCTGAAGTACACAGAGCGCTTGGACACAGAGTGTGCTGGTGTGGACAGGAGG 1106
DB 167 GlySerGly-----TipArgArgArgGly-----ArgThrArg 178
QY 1105 GCGCAGCTTCTTCTCTG-----ACAGCTCTCAATGCGCGGTGACGC 1061
DB 179 ArgProAlaAlaSerArgArgProAlaGlyGlyProAlaGlyArgProGlyGlySerArg 198
QY 1060 CCTCGCGCAACAGCTGCCACAGCTCTCAACTCCAGCTCTGAATGCCACCTTCTTCA 1001
DB 199 AsnArgArgAla----- 203
QY 1000 GTCTCAGAGCGCGACTGCGCGGGTTGAGTTCACCGACACCACTGCGCGCGCCCT 941
DB 204 -----GlyCysArgProAlaArgProValAlaArgAla 215
QY 940 GCACACAGGTTTCAGGGCATTCGACGCGCGCTAGTACGGGTGAGCCAGCTGCTTGG 881
DB 216 GlyPro-----GlyProGlyArgCysArgArgArgCysAlaTrpSerProCysPro- 233
QY 880 GTTTGTATCTCCATCAGCTCCATATCCGCTCGGGTCTCCACAGCTGCTGTACAGGA 821
DB 233 ----- 233
QY 820 AGTGGCGGGAACATGCGCGCGCTCCAGCGCTCTCTGTGAGGCGGTGGCCACGC 761
DB 234 -----ThrProArg-----AspArgAlaGlyAlaGlyAspArg 244
QY 760 TGGCAGCGCGCCCAATCCACAGCTAGTAGGGG-----CGCGCAGCCACGCGCAGT 710

```

Db 245 HisValProAlaGlyAspAlaArgAlaAlaGlyTrpProThrGlySerArgCysAlaArg 264
Qy 709 AGGGATCGAACCTGGGTGTTCTGCTCCACAGCGCTCCAGCTTCTGCTCCAGGTAGG 650
Db 265 ArgSerGlySerAlaGlyCysAlaArgPro-----ProArgArgSerCysValArg 281
Qy 649 CGCGGCGCTCGGGCCGATGTCATGTTGTCGATCGAAGATCGATCGCCAGAGAAC 590
Db 282 TrpAlaAlaArg-----ArgValAlaArgAspSerAlaAlaArgSer 295
Qy 589 GCGCGAGACCGGCGATCTGCGCGAGGCGAGGTGCTACTTGCCTCAGCTAGAGTCGG 530
Db 296 AspArgProValArgSerAlaLeuArgGlyLeuProAlaAlaArgSerArgProArg 315
Qy 529 CAACGCCACAGCGCTCTTGGACAGGTACAGAGCAAGCTGTGTCGATGACGTGTGGA 470
Db 316 HisArgArgProGlySer-ValProProValArgThrSerArgAla-----ThrProAr 333
Qy 469 ACGTGTGAATCATCGT-----GAGCGAGTAGGAGAGGTGATGAGCGTCG 425
Db 333 GArgTrpHisAlaArgTrpGlnAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 353
Qy 424 CGGTGCCCTCA-----GCGGCGCGCAAAATTTGCAAGCGCTCGCGCTCCACGA 380
Db 353 lathProAlaAlaProGlyValArgGlyTyrProArgGluTyrAla-ArgAlaProArg 372
Qy 379 CCGTGACATTTTCCAGCCCT-----TGGCTTTCGCTTCTTCTTTCGCGCA 335
Db 373 ProAlaLeuProGlyArgProAlaAlaArgProAlaArgProArgProCysArgArgPro 392
Qy 334 CTTGCGACAGGAGTGCACAGTCGACCA-----CGTAGATGACT 293
Db 393 AlaAspArgSerProGlyGlnGlyArgProGlySerSerAlaThrArgArgTrpArg 412
Qy 292 TGAATTCGCGAGTCGATGATATCAGCCATCATTCGACATCTCCCGCAGTCCACAC 233
Db 413 -----ProCysHisArg 416
Qy 232 CCAGTCAACCCAGATGAGTTTCGAGCGCTCGGCCAGC-----GGGCAG 188
Db 417 ArgAlaGluProAlaAlaAlaProAlaGlyTyrProGlyAlaAlaAlaArgSerAlaThr 436
Qy 187 CAAGGAGCGGCTGGGCGCGGTAGAACG---TCTCCAGGCGAGCAGCGTATCATCGC 131
Db 437 ArgArgThrHisAlaGlyArgArgGlySerAlaThrProSerGlyArgArgAspArg 456
Qy 130 CTTTCT 125
Db 457 ProAla 458

```

RESULT 9

```

US-09-252-991A-31760
; Sequence 31760, Application US/09252991A
; Patent No. 5551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31760
; LENGTH: 1706
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31760

```

Alignment Scores:

```

Pred. No.: 6,27e-15 Length: 1706
Score: 316.50 Matches: 193
Percent Similarity: 33.1% Conservative: 44
Best Local Similarity: 26.9% Mismatches: 249
Query Match: 8.7% Indels: 231
DB: 2 Gaps: 36

US-10-620-914-44 (1-1947) x US-09-252-991A-31760 (1-1706)
Qy 1846 CGTCGAAGCCCGGCTTCTGATCAGCTCGCGGTAGGCGCGGCTCAGGAGCGGAGCGCC 1787
Db 276 ArgArgGlnProAlaGlyAlaThrAlaArgLeuArgHisArgProGly---HisProAla 294
Qy 1786 AGATGACGATCGCCCGCGCAACCTGCTTGGCCAGGCACTCGGCCAGCTCGTGTGCCCA 1727
Db 295 ArgLeuArgArgGlnProAlaArgLeuArgSerGlyGluArgProAlaGluSerPro 314
Qy 1726 CGGCGCATATCCAGCCAGTCCACGTGT----- 1700
Db 315 ArgArgThrProGlyGlnLeuArgGlyArgArgProAspAlaLeuProArgLeuAlaGly 334
Qy 1699 ---CCATCAGATCAGCT-----TGGTGTAGGTGGCGCTTTGAGCTCCT---CCATGA 1652
Db 335 LysProGlyArgTyrProAlaGluPheAlaArgGlyAlaAlaProAlaProGlyPro--- 353
Qy 1651 AGAAGTTGTGGAGCGGTGCTCCACACGCGCTCTTTGAGGG-----TGG 1601
Db 354 -----LeuAlaAlaValProArgArgHisAlaAlaArgAlaAlaGluProTyr 369
Qy 1600 CGAAGCGCGCTCGCGCAGTAGTGGGCGAGTGTTCGCGCAGGAACTTCCCGGTGAGGC 1541
Db 370 ArgArgProAlaGlyGlyArg-----AlaAspArgThrAlaThr----- 383
Qy 1540 AGTTGTAGTAGAAGTAGTTCTGCTGCGCAGCGAGTTCTCGCGCAGCGCTCCATGG 1481
Db 384 AlaProAlaArgProArgThrAlaCysAlaAlaValAlaGlyProArgHis----- 401
Qy 1480 TGC CGCGATGTAGTTCTCAATGGGATGCGCTCGCGCTTGATCAGCGCTACTGTGTC 1421
Db 401 ----- 401
Qy 1420 CGGCGAGCGCGCGCGAACCAGCAGCGCTTGTGTGAAGACAGCAGGCTCAGCA--- 1364
Db 402 ---AlaArgArgGlnProArgThrLeuProGlyProArgProGlyAlaArgGly 420
Qy 1363 -----ACTTGACGAACAGCAGCAGCGCTTGGCGCTTCTTCACGA 1319
Db 421 ThrHisLeuAlaGlySerAlaGlnProAlaAlaProAlaAla---AlaArgProPro--- 438
Qy 1318 AGTGGATGAGCATGTTGCTGTCCACAGCGCGCTGCTCCTCATTTGTGGCGCGTGG 1259
Db 439 -----AlaAlaLeuProAlaAlaArgAlaArg--- 446
Qy 1258 CGAGGCGCTGACGCTTGTCCCGCAGTCCAGCAGCA-----CGGCCAGGCTCAGCA 1208
Db 447 ArgGlyArgArgGlnProProGlyArgProArgLeuArgProGly---Ala 465
Qy 1207 GCACCAGCAGCAGCTTGCCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1148
Db 466 ProValAlaAlaAlaAlaPro-----AlaThrGlyAlaAlaThrGluPro 481
Qy 1147 GCGCTTGGACCAAGATTTGTGCTGTTTGGCAGCAGGAGCGCGCGCTTCTTCTCGT 1088
Db 482 GlyAla-----GlyAspSerHisArgCysArg 491
Qy 1087 ACAGCTCTCAATGCGGGGTGACCGCTTCCGCGAAGCAGCTGCCACAGCTCTCAACT 1028
Db 492 SerGlyThrAspArgProGly---ArgProArgGlnProAlaAlaAlaArgAspProGly 510
Qy 1027 CCAGCTGCTGAATGGCACCTTCTTCAGCTCCAGAGCGCGGCTGCGCGGGTGTGAGT 968
Db 511 ProGlySer---TrpProGlyProSerAlaAlaArgAlaAlaArgGly---Asp 528

```

```
Qy 967 CCA-----CGGACA 959
Db 529 ProHisArgArgGluThrAlaLeuArgSerAlaAlaArgArgProArgArgAlaAsp 548
Qy 958 CCACCTGGCGGGCCCTCGCAGCAGAGGTTTCAGGGCATTTCAGCGCCGCTAGTCAGGG 899
Db 549 ProProGlyArgThrProAlaGlnProGlyGlnThrGluGlyAlaArgArg----- 565
Qy 898 TCAGCAGCGTCTCTGGGTTGATCTCCATCATCTCCATATCGGGCTCGGGTCTCTCCC 839
Db 565 ----- 565
Qy 838 ACGACTCGGTACAGGAAGTGGCGGGAACATGGCGCGCTCTCCACGCGCTCTCCGT 779
Db 566 -----LeuAlaArgArgThrAlaGlyThr-----AlaAlaProArg 578
Qy 778 GCAGGGCGTGGCCCAACCTGGCGAGCGGCGCAATCCACAGTGTAGGGGCGCGCAGCC 719
Db 579 HisArgArgGlnProGluGlnProAlaGlyProArgProAlaArgGlyAlaThrAla 598
Qy 718 ACGGCAGT-----AGGGATCGAACCTGGG 692
Db 599 ThrSerArgLeuProAlaLeuProTrpProArgArgProAlaArgSerHisArgProGly 618
Qy 691 TGTCTCTCCACACGCGCTCCAGCTTCTCTCCAGTAGGCGCGCGCTCGGGCGCA 632
Db 619 IleProGlyProAla-----AlaProAlaProGlyThrThrArgArgAspSerArgArg 637
Qy 631 TGTCAATGTTGTCTGATCGAAGATCGATCGCAGAAAGAACGGCGCGCACCGGCA--- 575
Db 638 -----ThrArgSerHisGlyAlaProHisAlaGln 647
Qy 574 -----TCTGGCGCAGGGCAGGTCTGACTTCGCTCAGCTAGTAGTGGCAACGC--- 524
Db 648 ProArgSerAlaAlaGlyLeuGluAlaAlaProGlyThrArgGlnSerGlyThrAlaVal 667
Qy 523 CCACGAGCGCTCTGGGACAGGTACGAGCAAGCTGTGTGATGACGTGTGGAAACGGTG 464
Db 668 GlnProGlyArgSerGlnGlyGlnGlyArgGlnProArgArgSerArgCysAlaAspAla 687
Qy 463 GAATCATCTGTGAGCGAGTAGGAGAGGTGTAGCGCTCGCGTGCCCT---CAGGGGCGC 407
Db 688 ArgAlaGluProArgThrAlaArgGlnGlyHisArgArgArgAlaProGlyGlnProGln 707
Qy 406 CAATTTGGCAAGCTCGGCTCCAGCAGCTCGACATCTTCCAGCCCTTGGCCTCGCCT 347
Db 708 AlaMetGlyArgArgHisProPro-----AspArgGln 718
Qy 346 TCTTTTGGCCACCTCGCAGCGAGTGGCAGCTGACACAGCTAGATGAGTGAAC 287
Db 719 AlaLeuAlaProAlaAlaArgValHisArgAlaGlyArgArgThrArgProAspGluPro 738
Qy 286 TCGCCAGCTCATGTAATCAGCCATCATATCGACATCTTCCCCAGTGCACCCAGGT 227
Db 739 AlaProGlyLeuAlaAspGluPro-----ArgProGly 749
Qy 226 CAACCCAGATGAGTTCGAGCGCTCGGCA----- 197
Db 750 GlnProArgAlaAlaAlaGlnGlyArgAlaValArgHisGlyAspLeuArgSerLeu 769
Qy 196 -----GGCGGCGAG 188
Db 770 AlaAspAlaGlyValArgProAlaAspProLeuSerArgAlaGlyHisGlyGlyGln 789
Qy 187 CAAAGG-----CAGCGGCTTGGGGCCCGCTAGA 161
Db 790 ArgArgArgGluAlaAspAlaAlaAspArgValLeuLeuGlnProArgGlyLysArg 809
Qy 160 AGCTCTCCAGCGGAGCAGCGTATCATCGCCCTTCTTCTCGCGCAACACATATGGCGCA 101
Db 810 GlyArgThrValArgArgSerAlaArgArgGlyArgArgArg-----GlyAla 826
Qy 100 GAACGGTCA-----GGTCATCTTCTATGCTGCTGAGCTTCTCCAGGAGAGT 47
```

```
Db 827 AlaArgSerLeuArgGlyHisLeuGluProProArgAspGlnGlyLeuProGlyProPro 846
Qy 46 TCTTCTTGGTGTAGCTCGCAGGCGCGCGCTCAC-----GACCGGACCCCA 2
Db 847 AlaThrGlyAlaGlnArgProAlaGlnHisHisProAlaAspProLeuPro 863

RESULT 10
US-09-252-991A-27341
; Sequence 27341, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27341
; LENGTH: 1073
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (803)
; OTHER INFORMATION: Identity of amino acid at the above locations are unknown.
US-09-252-991A-27341
```

```
Alignment Scores:
Pred. No.: 1.58e-14 Length: 1073
Score: 310.00 Matches: 243
Percent Similarity: 28.1% Conservative: 41
Best Local Similarity: 24.0% Mismatches: 288
Query Match: 8.6% Indels: 440
DB: 2 Gaps: 47

US-10-620-914-44 (1-1947) x US-09-252-991A-27341 (1-1073)
Qy 18 CGGCGGCGCTCGAGCTACAC-----CAAGAAGAACTTCTCCTCGGAGAGCTCAA 68
Db 92 ArgCysSerCysThrLeuGlyAlaSerIleArgGluGluArgGlnProAspGluArgLys 111
Qy 69 GCTCAGCAGCATGAAGGATGACCTGACCCGTTCTGCGCCATATGTGTTCGGCAGCAAGAA 128
Db 112 AsnAlaArgHis-----ValProArgArg 120
Qy 129 GGGCGATGATCAGCTGCTGCTGCGTGGAGAGCTTCTACGGGCCCCCAGGCGCTGCTTGC 188
Db 121 ArgArgProGlyProCysArgProGlyProLeuHisAlaArgLeuProArgValArg 140
Qy 189 TGCCCGCTGCGCGAGCGCTCGAACCTCATCTGGGTTGACCTGCGGTGTGTCGACTGGGGA 248
Db 141 ProProArgGlnArgLeuAlaGlu----- 148
Qy 249 GAATGTCCATATGATGCTGCTGATTACATCGACCTGCGAAGTTCAAGTCCATCTAGTGT 308
Db 149 -----HisArgArgArgArgArgArgLeuTyrLeuArgHis 160
Qy 309 CGACTGTGCCATCGCTGCTGCGAGGTGGCCAGAA----- 344
Db 161 ArgArgGlnProLeuProArgArgGlyGlyArgHisValHisGlnHisArgProGly 180
Qy 345 -----GAAGCGAAGGCGCAAGGCTGGAAGATGTCACGCT----- 380
Db 181 AlaArgGlySerHisArgGlyArgAlaAspProProAlaGlyLeuPheGlnSer 200
Qy 381 -----CGTGGAGGC 389
```

201 LeuLeuArgHisGlyGlnProAlaArgHisArgThrLeuProGlnAlaArgArgAlaGly 220
Qy CGAGCTTGGCAATTGGCCGCC-----TGAGG 419
Db ProArgArgPro-----ArgProArgValProHisArgPheHisArgArgGly 238
Qy CACCGCAGCGTTCATCACCCTTCTCTA----- 446
Db HisArgAspProProHisAlaLeuLeuProGluLeuProArgGlnAlaArgGlnGluAla 258
Qy CTCGCTCAGGATGATCCACCGTTCCA----- 473
Db ArgHisHisAlaAspGlnArgLeuProArgLeuAspLeuProArgHisValAlaGlyArg 278
Qy CAACGTCATCGACCA----- 488
Db GlnGluArgArgProAlaGlyArgValArgLeuProArgAlaHisProProCys 298
Qy ---GGCTTGTCTGACCT-----GTCCCAAGCGCCT 518
Db LeuSerLeuLeuLeuProArgSerGlyArgSerGlyArgValProArgTyrPro 318
Qy GGTGGCGGTTGCCGACTT----- 536
Db GlyGlyGlyValArgThrGlnAspProArgThrGlyArgArgProGlyGlyGlyValHis 338
Qy ---CTACGTGAGCGGCAAGTACGACCTGCGCCCTCGCCCA 572
Db LeuArgAlaGlyValArgLeuArgArgAspArgProAlaArgGlyLeuProProAla 358
Qy GATGCCCTGTGGCGCGTTCCTT---CTGGCGATCGATCTTCGACATCGACACATTGA 629
Db AspValGlySerValProAlaLeuArgArgAlaValHisLeuArgArgSerGlyAspLeu 378
Qy CATCGGCCCGCGAGCGCG----- 647
Db LeuArgProProArgProLeuLeuArgGlnProGlyGlyValArgArgThrAlaGlyHis 398
Qy ---CGCTACTCGAGCA----- 662
Db HisProHisArgGlnGlyProHisLeuArgLeuProAlaAlaGlyArgValHisLeuLeu 418
Qy --- 662
Db ProAlaHisLeuGlyAspArgAlaGlyGlnGlyProLeuLeuGlnProTyrPhe 438
Qy ---GAAGCTGGAGCGGTGGGAGCAGACACCCAGGTTCCGATCCC 707
Db HisLeuLeuArgProProGlyGlyLeuArgGlyAlaGluGluHisArgAspHisArg 458
Qy CTACGTCCGCGTCT---GCGCGCCCTTACTACCTGTGGATTGGCCG 752
Db AlaArgGlyLeuAlaArgProArgArgArgGlyArgProLeuLeuArg-----GlyAla 476
Qy CTCTCCAGCGTTGGCCACGCGCTGCA----- 779
Db ProAlaHisSer---ProArgProAlaHisArgArgArgAlaArgAspAlaLeuHis 495
Qy ---CGAGGCGCGTGGAGGAGCCCGCGCCCTGTTCCCGCC 815
Db GlyLeuCysArgValArgArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 515
Qy CACCTCTCTGACGAGTGTGGGAGGAGCCCGAGCGGATAT-----GGA 863
Db HisArgArgVal-----GlyProProAlaGlyAlaGluAlaArgProAla 530
Qy GTGATGAGATCAACCCCAAGGACAC-----GGTGTGACCTGTAGCTAGCGCGG 914
Db GlySerSerAspArgProProGluArgAspValAlaAlaAlaAspProHisProArgThr 550
Qy CTGCAATGCCCTGAACTGTGGTGGAGGGGCGCGCGCGTGTGTGGTGGACTGCA 974
Db GlyArgTyrArgGlyProGlyAlaAlaArgGluHisArgGlyAsnArgGlyGlySerCys 570

Qy CCCGCGCAGTCGCGCGTCTCT---GGAGCTGAAGAAGGTGGCCATTGAGCAGCTGGAGTT 1031
Db ProArgArgSerProValSerAlaGlyAlaProProAlaProHisProAsnHisGlnGlu 590
Qy TGAGGACGTGGCAGCTGTCGGCAGGCGGTGCACCCCGCATTTGAGGAGCTGTACGA 1091
Db AsnProHisArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 610
Qy GAAGAAGCTGGCGCCCTCTCTGTCGCAACAGCAGCCACCACTTCTGTGTCGAAGCCCTCTG 1151
Db GlyLeuGlyAspAlaLeuHisGlyProAspArgProGlyPheLeuProArgSer----- 628
Qy GTACTTCCAGCAGCGCTGTACTACAGGCGCGCATGGCAAGCTGTGTGGTGTGTGCA 1211
Db ---ValGluProValLeuValAlaLeuArgAla-----GlyProAla 641
Qy GTCTCTGCGCGTGT-----GCTGGGACTGGGCAAGCCGTCAC----- 1250
Db ValProAlaAspGlyHisAspAspLeuArgAlaGlyHisArgLeuProArgGlnArg 661
Qy ---GCGCTCGCCACGCGCCGCCACCAATGAGGAGCAGCGCCCTCTGTGTG 1295
Db ArgLeuProLeuGlyAlaGlnArgLeuArgLeuGlyGly-----ProGlyVal 679
Qy ---GGACAGCAACATGCT-----CATCCACTTCGTGAA 1325
Db MetAspValLeuGlyGlnGlnArgThrValAspAlaPheGlyLeuHisProValArgGln 699
Qy ---CGGCGCCACGCGCTGTGTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1364
Db ---CTTCAACAGCGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1397
Qy ---GlyValGlnArgGlyAlaAlaProGlyGlnVal 738
Qy GTGGTTCGCGCGCGGTGCGCGCGAGCAGTACGC---GCTGATCAAGCGGCGCGCAT 1454
Db AlaAlaGluProArgArgAlaAlaGlnAlaAlaArgGlyAlaArgProGlyArgArg 758
Qy ---CATTGAGAACTACATCGCGC 1478
Db ProAlaLeuArgLeuGluProTyrProLeuArgGlnArgLeuGlnProGlyLeuHisArgPro 778
Qy ---CACCATGGAGCGGT 1493
Db ValLeuAlaGlyProAspGlyArgAlaGlySerAspGlyLeuTyrHisHisGlyAspArg 798
Qy GCGCGAGAACTCGCACGTGCGCAAGCAGAACTACTTCTACTACAACTGCTCACCAGCA 1553
Db AlaGlyLeuLeu-----ArgArg 805
Qy GTCTCTGCGCGCACTGCGCCACCTTACTCTCTGCGGAGCGCGC----- 1595
Db GlyTyrAlaGlnArgAlaProArgHisProAlaArgGlyGlyAspLeuArgAlaAspArg 825
Qy ---CTTCGCCACCTCAAGAG---TGGCGTGTGT---GGACAACTGAC 1634
Db ArgArgLeuGlnHisLeuArgHisProArgArgAlaGlyArgGlyThrGlyArgGlyAsn 845
Qy CGTCTCCAC----- 1643
Db ArgArgHisProAspLeuArgProTyrProLeuGlnHisLeuArgProArgArgArg 865
Qy ---CAACTTCTCATGGAGGAGCT 1664
Db ArgAspAlaArgArgProGlyArgArgLeuArgProValHisProLeuHisGlnHisGly 885
Qy CAAAGCGCGCACCTTACACCA-----GGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1712
Db HisLeuGluHisGlyHisGlnProCysArgGlyGlySerGlyGlnGlyArgArgAlaAla 905

QY 1713 GCTGATATGCCGTGGCCACGAGCTGGCCGAGTGCT-----GGCCAGCAGGT 1763
Db 906 GlyAlaValArgArgGlyProPheAlaProArgHisAlaAspArgLeuGlyGlyAlaGly 925
QY 1764 TGCCTGGCGGCGCATCTG----- 1781
Db 926 LeuGlyGlyGluHisArgProAlaAlaLeuArgProGlyGlyAlaHisCysArgGly 945
QY 1782 -----CATCTGGCGCTCCGCTCCCT 1802
Db 946 ThrValLeuAspProAlaValAspLeuArgHisGlyPheHisAspAlaLeuCysAlaAsp 965
QY 1803 CAGCCGCTTACCGCGAGCTGATCCAGAA----- 1832
Db 966 ValProGlyLeuArgProProAlaProArgArgProAlaSerThrAlaLeuProAspAla 985
QY 1833 -----GGCGGCTTCGACGTGGCGCTG----- 1853
Db 986 AlaGlyArgProProGlyGlnSerLeuGlyAlaValArgAlaAlaCysAlaGlyArg 1005
QY 1854 -----CATCCGCGCGCCACTCAGGCTACATGACCGCTCAACAT 1895
Db 1006 HisLeuProValArgArgHisProArgAlaAspGlyLeuGlyLeuArgArgGlnAsp 1025
QY 1896 GTACAGCTCTTCTACATGCGCCGCGGAGGCGC 1931
Db 1026 ArgArgArgArgGlyAlaGlyProGlyArgGlyArg 1037

RESULT 11

US-09-252-991A-23774
; Sequence 23774, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23774
; LENGTH: 798
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23774

Alignment Scores:

Pred. No.: 1.66e-14 Length: 798
Score: 309.00 Matches: 213
Percent Similarity: 30.4% Conservative: 33
Best Local Similarity: 26.3% Mismatches: 255
Query Match: 8.5% Indels: 309
DB: 2 Gaps: 47

US-10-620-914-44 (1-1947) x US-09-252-991A-23774 (1-798)

QY 1933 TGGCGCCCTTCGCGCGG-----CCATGTAGAGGAGC 1901
Db 16 TrpProMetAlaGlyGlySerMetAlaCysValAlaArgTrpThrProCysArg----- 33
QY 1900 TGTATCATGTTGACGGCGGTCATAGCCCTGAGTCGCGCGCGGATGACGCGCAGCTCGA 1841
Db 34 -----ArgGlySerCys-----ArgTrpLeuAlaIleCysArgLeuArgSer 47
QY 1840 ACCCGCCTTCTGGATCATGCTCGGCTAGGCGGCGCTG-----AGG 1800
Db 48 ValProArgSerGlyArgAlaSerArgSerThrTrpCysThrAlaProProVal 67
QY 1799 GAGCGGAGCGGCAGATGACGATGCGCGCGCGCAACC-----TGCTTGCGC 1752

Db 68 SerThrThrArgLeuAlaThrValProProThrSerThrAlaCysAlaMetCys-TrpAl 87
QY 1751 AGGCACTGCGCCAGCT----- 1736
Db 87 aGlyCysArgProAlaAlaSerGlyProGlyAlaCysCysSerSerProAlaProGlySe 107
QY 1735 -----CGTTGGCCACGGGCATA 1719
Db 107 rThrArgArgProMetAlaAlaGlySerThrArgSerArgArgArgPheProArgSerIl 127
QY 1718 TCCA-----GCCAGTCCAGTGGTCCATCAGAAATCATTGGTGT----- 1679
Db 127 eProAspGlySerCysSerThrProSerArgSerArgTrpThrAlaGlyTyrArgLeuPr 147
QY 1678 -AGGTGGCGCTTTGAGCTCTCTCATGAGAGTGTGGTGGAGACGGTCAGGTTCTCCACC 1620
Db 147 oAlaCysAlaTrpProAlaSerThrValArgGly-----GlySerGlyCysSerTh 164
QY 1619 ACGC-----CACTCTTTGAGGTGG-----CGAAGGCGC----- 1592
Db 164 rArgCysValArgGlyThrGlyTrpSerAlaSerArgProCysMetProThrGlySerTh 184
QY 1591 -----CCTCGCCAGGTAGGTGGCGAGTGTTCGCGCAGGAACCTTCGCGGTGAGGAG 1539
Db 184 rProThrMetArgProGlyCysTrpProSerCys-----CysAlaProMetPr 200
QY 1538 TTGTAGTAGAAGTAGTCTCTGCTTCGCGCAGTTCCTCCGCGCAGCGCTCATG----- 1481
Db 200 oAlaAlaArgArgTrpArgThrAlaThrSerValSerThr-----ThrArgProTrpAr 218
QY 1480 -TGCGCGCGA-----TGATGTTCTCAATGG-----GGATGCGCTCGCCTTGATC 1437
Db 218 gCysThrArgTrpSerThrCysAlaSerAlaTrpAlaSerAlaAsnGlyProThrAs 238
QY 1436 AGCCGCTACTGCTTGGCCGCGCACCGCCGCGCGCAACACACGACGCGCTTTGTGAGAGC 1377
Db 238 nileArgTyrAlaAlaProAlaAlaSerValAlaAlaThrAlaAlaProGly---ArgSe 257
QY 1376 ACAGGCTCAGCACTTGACGACGACGACGACGCGCTTGGCGCGCTTCTTCACGAG 1317
Db 257 rAlaGlyCysArg-----AlaIleProAlaThr---AlaArgAlaThrArg-- 271
QY 1316 TGGATGAGCATGTTGCTTCCACAGACGCGCTCTCTCCATTG-----TGGGC 1266
Db 272 -----ArgCysSerGlyThrProAspGlyCysGlnProValValGlyAlaSerLeuAl 289
QY 1265 GCGTTGGCGAGCGCTTGACGCTTTCGCCAGTCCAGCACGACGCGCAGGCGACTGCGAGC 1206
Db 289 aAlaTrpArgGlyAla-----ValProCysGlyLeuLeuArgProGlyValArgPr 306
QY 1205 ACCCAGC----- 1199
Db 306 oProGlnProPheGlyLeuArgHisArgLeuAspProProArgArgArgArgLeuProVa 326
QY 1198 -----ACAGCTTGGCCATGCGCGCTCGCTG-----AGTACAGCGCGCTGC 1161
Db 326 lProArgAlaAlaProCysAlaAlaGlyArgHisProValAlaAspGluProGlyArgAl 346
QY 1160 TGAAGTATACAGAGCGCTTGGACACGAGTGTGGCTGG-----TTTGGCACAGG 1110
Db 346 a---ThrTyrArgAlaArgArgGileArgAlaGlyLeuProAspAlaLeuCysArgGl 365
QY 1109 AAGGGCGCAGCTTCTTCGTACAGCTCTCTCAATGCGCGGTGACGCGCTTCGCGCAAC 1050
Db 365 yAsnAlaProAlaArgPro---AlaArgProGlnThrValAlaAlaArgLeuArgTh 384
QY 1049 AGCTGCCACAGCTCTCTCAAACTCCAGCTGTGTAATGGCCACCTTCTTCAGCTCCAGAAGC 990
Db 384 rGluProGlyArg---ArgArgGlnArgGlyGlyTrpProGlyValAlaGlyArgGlyPh 403
QY 989 GCCGACTCGCGGGGTTCAGTCCACCGACACCACT----- 953

Db 403 eProAlaGlnArgArgIleGlyValGlyLysProProAlaGlyLeuAlaArgThrGlyLe 423
Qy 952 -----GGCGGCCCCCTGCACAGCAGGTTTCAGGGCATTCGACGCGCCCTAGTCAGG 900
Db 423 uArgAlaSerArgArgLeuAlaProGlyGlyArgGlyAlaProArgArg----- 440
Qy 899 GTCAGCACCGTGCTTGGGGTTGATCTCCATACCTCCATATCCGGTCCGGGTCTCTCC 840
Db 441 -----AsnHisArgThrProAlaArgLeuProG1 450
Qy 839 CACGACTCGGTGTACAGGAGGTGGCG-----GGAACATGGCGCGCTCCACG 789
Db 450 ySerAlaCysTyArgGlyAlaArgSerGlyProThrGlyAlaValAlaAlaProPr 470
Qy 788 CGTCTCTGTGTCAGGGCGGTGGCCAACTGGGCGGCGCAATCCACAGTAGTAGGG 729
Db 470 oAlaPro-----ProAspArgG1 476
Qy 728 GCGGCGACCCAGCGCAGTAGGGATCGAACCTGGGTGTTCTGCTCCACACCGCTCC 669
Db 476 yPheProAlaCysLeuArgArgAlaAlaAlaTyLeuProThrAlaMetProGlyAlaPr 496
Qy 668 AGCTTCTGCTCCA-----GCTAGCGCGCGCTCGGGCGCGATGTCATGTTCTCGCATG 615
Db 496 oGlyGlnAlaProValAlaGlyArgThrAlaAlaGlyGlyThrArgProAlaProArgLe 516
Qy 614 TCGAAGATCGATCCGCCAGA----- 596
Db 516 uLeuArgProGlyAlaLeuLeuHisAspLeuProAlaPheHisArgGlyHisSerGlyAl 536
Qy 595 -----AGAAACGCGCGCGACGAGCATCTGG 570
Db 536 aValProAlaGlnArgGlnAspLeuSerPhePheGlnTyArgAlaArgAlaAlaG1 556
Qy 569 CGCAGGCGCAGGTGCTACTTCCCGCTCA-----CGTAGAGTGGCA 528
Db 556 nThrGlyGlyLeuArgValSerArgGlyLeuHisValArgArgValArgProCysArgG1 576
Qy 527 ACGCCCA-----CCAGCGCGCTTGGGACAGGTACGACGAGCGCTGTCGATGACG 477
Db 576 nTyProPheArgArgProAlaValAlaGlyAlaGlyLeu----- 589
Qy 476 TTGTGGAAACGTGGATCATCTGTCAGCGAGTAGGAGAGGTGATGAGCGTGGCGGTGCC 417
Db 589 ----- 589
Qy 416 TCAGGGGCGCAATTGCGAAGCGTGGCTCCACGACCTGGACATTCCTCCAGCCCTTG 357
Db 590 -----LeuProValGlyAlaArgArgPro----- 597
Qy 356 GCCTTCGCTTCTTCTTGGCCACCTCGCACAGCGAGTGGCACAGGTGACCAAGTAGATG 297
Db 598 -----AlaAlaArgPro----- 601
Qy 296 GACTTGAACCTTCGCGAGTGTATATCAGCCATATATCGACATTCCTCCCGAGTGCCA 237
Db 602 -----AlaProGly-----ArgArgGlyGlyArgArgHi 611
Qy 236 CCACCCAGGTCAA-----CCAGATGAGTTTCAGCGCTCGCGCCAGCGGGCGAGCAA----- 185
Db 611 sArgProGlyGlnProProGluHisAlaAlaGlyAlaArgProValGlyGlnArgAlaG1 631
Qy 184 -----AGCAGCGGG-----CCTGGGGCGGTAGAGC----- 158
Db 631 yLeuGlyAlaGlnArgLeuAlaAlaAlaGlyAlaArgArgThrValLeuProLeuAl 651
Qy 157 -----TCTCCAGGGGAGCAGCGTGTATCCTCCCTTCTGCTCGCGCAACCAATATGCGCG 102
Db 651 aGlyGlyProGly-----AlaAlaCysAlaGluAlaThrGlyAsnAl 665
Qy 101 AGAACGTGAGTTCATCTTCATGCTGCTGAGCTTTCAGCTTTCAGGAGAGAGTTCCTTC 42
Db 665 aGlyAlaAlaArgArgPro-----ValProArgArgLeuSerAl 678

Qy 41 TTGGTGTAGCTCGCAGCGCCGCC 19
Db 678 a-ArgAlaAlaArgArgProAla 685
RESULT 12
US-09-252-991A-17508
; Sequence 17508, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 17508
; LENGTH: 1064
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17508

Alignment Scores:
Pred. No.: 2,21e-14 Length: 1064
Score: 308.00 Matches: 215
Percent Similarity: 33.9% Conservative: 43
Best Local Similarity: 28.2% Mismatches: 267
Query Match: 8.5% Indels: 237
DB: 2 Gaps: 44

US-10-620-914-44 (1-1947) x US-09-252-991A-17508 (1-1064)

Qy 149 GCCTGGAGAGCTTCTAGCGGCCCGCCGCTGCTTGTGTCGCC----- 193
Db 9 AlaTyrSerAlaAlaAlaTyrProTyrProAlaSerSerSerProCysSerSerCys 28
Qy 194 ---GCCTGGCGGAGC-----GCTCGAACCTCATCTGGGTGACCTGGGTG 235
Db 29 CysAlaTyrProSerAlaProGlyProAlaThrAlaCysSerCysCysSer 48
Qy 236 GTGCGACTGGGAGATGTCGATATGATGCTGATGCTGACCTGCGGAGTTCAGT 295
Db 49 AlaAlaCysSer-----CysSerAlaTyr-----AlaCysSer 59
Qy 296 CCAATCTACGTGTGCGACCTGTGCGAGTGTGCGGAGTGGCCAGAGGCGAAGG 355
Db 60 AlaSerThrSerGlyAlaCysThrAlaThrCysAlaArg----- 72
Qy 356 CCAAGGGCTGGAAGATGTCAGGTGCG-----TGAGGCGCGAGCTGTCCTCGC 409
Db 73 -----GlyArgAlaSerSerSerAlaTyr-----SerAlaGlnArg 85
Qy 410 CCCTCAGGCGCACCGCGACGCTCATCA-----CCTTCTCTACTCGCTCAGCATG 463
Db 86 ProProArgPheProArgProCysSerGlyProValSerProLeuHisProAlaSerLeu 105
Qy 464 CACCGTTCACCAACAGTCATCGACCGCTTGTCTACCTGTCTCCCAAGA----- 512
Db 106 ProArgHisAspLeuGluSerArgArgLeuArg-LeuProArgHisArgLeuHisArgTy 125
Qy 513 -CGGCTGTGGCGGTGCGGACTTCTAGTCAGCGGCAAGTACGACCTCCCTCGGCC 571
Db 125 xArgSerProAlaGlnCysArgLeuArgAspArgArgLeuHisProCysArgArgPr 145
Qy 572 AGATGCCCTGGTCCGCCGCTTCTTCTGGCGATCGATCTTCGACATCGACAACTTGACA 631
Db 145 oThrGlyLysHisLeuLeuArgLeuGlyArgThrProLeuCysArgAlaArgHis----- 163

[illegible]

RESULT 13

US-09-252-991A-20408

; Sequence 20408, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 20408

; LENGTH: 957

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-20408

Alignment Scores:		
Pred. No.:	2,31e-14	Length:
Score:	307.50	Matches:
Percent Similarity:	34.6%	Conservative:
		32
		957

Db 441 aGlyArgGlyGlnProProAspAlaGlyAspProValAlaAlaGlyLeuSerArgAspAs 461
 Qy 524 CCCACGCGCGTCTTGGGACAGGTACGAGCAAGCCTGGTCGATGACGCTTGTGGAACGGT 465
 Db 461 pProGlnProThrGly-----ProGlyLeuProGlnArgGlyAspPr 476
 Qy 464 GGAATCATCGTGAGGAGTAGAGAGAGGTGATGAGCGTCCGCGGTCCCTCAGGGGGCGCA 405
 Db 476 oProArgPheSerAlaProArg-----ProAlaProGluHisArgPr 491
 Qy 404 AATTGGCAAGCGTCGCGCTCCACGA---CCTGGACATTCTTCCAGCCCTTGGCCTTCGCC 348
 Db 491 oLeuGlyHisGlnProAlaAlaArgGlnProGlyGlyThrSerArgAlaPro---Pr 510
 Qy 347 TTCTTCTTGGCCACT-----CGCACGAGTGCGC----- 317
 Db 510 oAlaThrHisProAlaGlyArgGlyAlaTyArgGlyArgGlyThrAlaGlyAlaLe 530
 Qy 316 -----ACAGTGCACACGTAGATGACTTGAATCTTGCAGGTCCGATGTAATCA 267
 Db 530 uProProAlaThrGlyGlnArgArgGlyArgAlaValAlaSerValGlyProAlaGlyAl 550
 Qy 266 GCCATCATATCGACATCTCC-----CAGTGCAC----- 236
 Db 550 aAlaProGlyGlnGlyAlaProGlyGlyArgArgGlyLeuHisArgArgProAspLeuPr 570
 Qy 235 -----CACCCAGGTCAACCCAGA-----TG 216
 Db 570 oAlaIleAlaGluLeuHisProAlaLeuProArgHisArgAlaGlyGlyProHisGlyAr 590
 Qy 215 AGTTCGAGCCTCGCCAGCGCGGAG---CAAAGCAGCGGCTGG----- 170
 Db 590 gGlyGlnArgArgGlyAlaGlyGlnGlyArgArgArgProGlyProGlyLeuAr 610
 Qy 169 -----GCCGTAGAGCTCTCCAGCGGAGCGGTGA 138
 Db 610 gProGlyArgGlyArgArgThrGlyArgArgAlaArgHisProProAlaProArgProAs 630
 Qy 137 TCAT---CGCCCTTCTTCTGCTCCGAAACACATATATGGCGCAGACGGTCAGGTCTCCTTC 81
 Db 630 pHisProArgGlnProCysProArgArgThr-----ProArgArgGly----- 644
 Qy 80 ATGCTGTGAGTGTAGCTTCTCCAGGAGGAAGTCTTCTTGTGTAGTCTCGCAGCGCGG 21
 Db 645 -----AlaAspAlaArgAlaAlaGlnArgValAlaGlyValaAspProGlnGlnHi 661
 Qy 20 CCCTCAGACCCGACC 5
 Db 661 sArgHisGlyProThr 666

RESULT 15

US-09-252-991A-26099
 ; Sequence 26099, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252.991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 26099
 ; LENGTH: 863
 ; TYPE: PRT
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-26099

Alignment Scores:

Pred. No.: 5.65e-14 Length: 863
 Score: 302.00 Matches: 227
 Percent Similarity: 29.8% Conservative: 29
 Best Local Similarity: 26.5% Mismatches: 252
 Query Match: 8.3% Gaps: 350
 DB: 2
 US-10-620-914-44 (1-1947) x US-09-252-991A-26099 (1-863)
 Qy 1939 CTTCTTGGCGCCCTTCGGCGGCCATGTAGA----- 1907
 Db 48 ProSerGlnArgArgArgGlyAspProProArgLeuArgLeuProLeuArgGluArgArg 67
 Qy 1906 ---AGGAGCTGTACATTTGACCGCGTCCATGTAGCCCTGAGTCGCGCGGATGACG 1850
 Db 68 IleArgProCysLeuArgSerArgArgAspCysPhe-----HisArgProGlyCysGly 85
 Qy 1849 GCAGCTCGAAGCCCGCTTCTGGATCA-----GCTCGGGCT 1814
 Db 86 SerHis-----ProProAspGlyGlnGlnAlaProValGluAspArgHisAlaArgArg 103
 Qy 1813 AGG-----GCGGGC 1805
 Db 104 ArgGlyAlaValHisProArgLeuArgArgArgThrGlyArgArgAspProAlaAla 123
 Qy 1804 TGAGGGAGG-----CGGAGCGCCAGATGACGATCGCCCG 1769
 Db 124 ArgGlyArgThrHisArgLeuProThrAspAspGlnGlyGlnArgArgArgArgPro 143
 Qy 1768 GCGCACTCTGTCGCGCAGCCTCGCGCAGCTCGTTGGCCA----- 1727
 Db 144 ArgHisAlaProGlyGlyThrArgProThrAlaGlyProValAlaHisArgThrLeu 163
 Qy 1726 -----CGGGCATATCCA 1715
 Db 164 GlySerAlaGlyArgLeuArgGlnArgArgThrAspProGlyGluGlyArgAlaAlaPro 183
 Qy 1714 GCCAGTCACCGTGTCCATCAGAAATCCTTGGTGTAGGTGCGGCTTGTAGCTCTCTCCA 1655
 Db 184 Ala-----ProCysGlyAspProGlyValArg-----ArgProPro 195
 Qy 1654 TGAAGAAGTTGGTGAGACCGTTCAGTTGTCCACCGCCACTCTTGAGGTGGGAGG 1595
 Db 196 -----ArgGlnProAlaLeuProArgArgThrArgLeuLeuGlyAlaAla 210
 Qy 1594 CCGCTCGCAGTGTGCGGCGAGTGTGCGCGAGAACTTCCCGTGGAGGAGTGT 1535
 Db 211 ProProGluGlyHisArgGlySerSerLeuProGly----- 223
 Qy 1534 AGTAGAAGTAGTTCTGCTTCGCGACGTCGAGTTCTCCGCCACCGCTCCATGTCGCGC 1475
 Db 224 -----AlaArgCysArgProAlaSerArgHisGlyArgGlyGlyGluGly 239
 Qy 1474 CGATGTAGTTCTCAATGGGATGCGCTCCGCTTGTATCAGCGGTACTGCTTGGCCCGCA 1415
 Db 240 ArgGlyPheGlyGlnLeuArgGlyArg----- 248
 Qy 1414 CGCGCGCGCGAACCACAGCACCG----- 1391
 Db 249 ArgHisArgArgValSerSerArgArgGlyTrpProLeuLeuProGlyAspGluTyr 268
 Qy 1390 ---CCTTGTGTAAGAGCACCGGCTCAGAACTTTCAGCAAGCAGCCACCGCGCTTGG 1334
 Db 269 ProProAlaGlyArgThrProGlyHisArgThrGlyHisArgProGlyProGlyLeu 288
 Qy 1333 GCCGCTTCTCAGAAAGTGA-----TGAGCATGTTGTGTCCACAGACGGCGCTGCT 1280
 Db 289 AlaAspProArgArgGlyProSerAlaAlaAlaAlaGlyThrGlyThrAla 308
 Qy 1279 -----CCTCATTTGCGCGGTTGGCGAGCGCTTGTGACGCTTGTG 1238
 Db 309 ArgAlaCysHisArgGlyProProLeu-----ArgArgGly-----ArgArgSer 323

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 14, 2006, 01:47:31 ; Search time 3 Seconds
(without alignments)
3612.968 Million cell updates/sec

Title: US-10-620-914-44
Perfect score: 3619
Sequence: 1 atggggctgggtcgtgacgg.....gcgccaagaaggacaactaa 1947

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 161667 seqs, 27834885 residues

Total number of hits satisfying chosen parameters: 323334

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh
-Q=/abs/ABSSWEB.spool/US10620914/runat_13032006_101939_5686/app_query.fasta.1
-DB=Published Applications AA New -QFMT=fastan -SUFFIX=n2p.rapbn -MINMATCH=0.1
-LOOPCL=0 -LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -HOST=abss05h
-USR=US10620914 @CGN 1.1.8 @runat_13032006_101939_5686 -NCPU=6 -ICPU=3
-NO MMAP -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WAE TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications AA New:

1: /cgn2_6/ptodata/1/pubppa/US08_NEW_PUB.pcp:*
2: /cgn2_6/ptodata/1/pubppa/US06_NEW_PUB.pcp:*
3: /cgn2_6/ptodata/1/pubppa/US07_NEW_PUB.pcp:*
4: /cgn2_6/ptodata/1/pubppa/PTCT_NEW_PUB.pcp:*
5: /cgn2_6/ptodata/1/pubppa/US09_NEW_PUB.pcp:*
6: /cgn2_6/ptodata/1/pubppa/US10_NEW_PUB.pcp:*
7: /cgn2_6/ptodata/1/pubppa/US11_NEW_PUB.pcp:*
8: /cgn2_6/ptodata/1/pubppa/US60_NEW_PUB.pcp:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	ID	Description
C 1	294.5	8.1	386	7	US-11-096-568A-21828
2	279.5	7.7	386	7	US-11-096-568A-21828
3	275	7.6	1886	6	US-10-515-868-8
C 4	270	7.5	495	7	US-11-182-016-31
5	267.5	7.4	428	7	US-11-096-568A-19503
6	267.5	7.4	493	7	US-11-096-568A-11657
7	267.5	7.4	544	7	US-11-096-568A-19501
8	267	7.4	1076	6	US-10-131-826A-219
9	267	7.4	1076	6	US-10-973-115B-219

10	262.5	7.3	1742	7	US-11-182-016-23	Sequence 23, Appl
11	262	7.2	413	7	US-11-096-568A-20771	Sequence 20771, A
12	261	7.2	1356	6	US-10-894-592-3	Sequence 3, Appl
C 13	258.5	7.1	766	7	US-11-096-568A-19864	Sequence 19864, A
14	257.5	7.1	1717	7	US-11-182-016-20	Sequence 20, Appl
15	256	7.1	1652	7	US-11-241-631-1	Sequence 1, Appl
16	255	7.0	375	7	US-11-096-568A-23618	Sequence 23618, A
C 17	254.5	7.0	459	7	US-11-096-568A-21887	Sequence 21887, A
18	253	7.0	2508	7	US-11-241-631-7	Sequence 7, Appl
19	253	7.0	2544	7	US-11-241-631-3	Sequence 3, Appl
20	253	7.0	2601	7	US-11-241-631-9	Sequence 9, Appl
21	251.5	6.9	1128	7	US-11-241-631-11	Sequence 11, Appl
22	250.5	6.9	788	7	US-11-096-568A-14593	Sequence 14593, A
C 23	249.5	6.9	493	7	US-11-096-568A-11657	Sequence 11657, A
C 24	248.5	6.9	413	7	US-11-096-568A-20771	Sequence 20771, A
C 25	247	6.8	428	7	US-11-096-568A-19503	Sequence 19503, A
C 26	247	6.8	544	7	US-11-096-568A-19501	Sequence 19501, A
27	246.5	6.8	1917	7	US-11-241-631-5	Sequence 5, Appl
C 28	242.5	6.7	373	7	US-11-096-568A-24110	Sequence 24110, A
C 29	240.5	6.6	369	7	US-11-096-568A-26217	Sequence 26217, A
30	239	6.6	558	7	US-11-096-568A-20255	Sequence 20255, A
C 31	238	6.6	449	7	US-11-096-568A-20255	Sequence 20255, A
32	238	6.6	575	6	US-10-980-388-46	Sequence 46, Appl
33	236	6.5	382	7	US-11-096-568A-26984	Sequence 26984, A
34	235.5	6.5	544	6	US-10-980-388-40	Sequence 40, Appl
35	235	6.5	442	7	US-11-096-568A-19628	Sequence 19628, A
36	234.5	6.5	535	7	US-11-096-568A-25889	Sequence 25889, A
C 37	233.5	6.5	512	7	US-11-096-568A-25890	Sequence 25890, A
38	232.5	6.4	325	7	US-11-096-568A-24112	Sequence 24112, A
C 39	232.5	6.4	459	7	US-11-096-568A-21887	Sequence 21887, A
40	232.5	6.4	579	7	US-11-096-568A-24333	Sequence 24333, A
41	232	6.4	369	7	US-11-096-568A-24111	Sequence 24111, A
42	232	6.4	373	7	US-11-096-568A-24110	Sequence 24110, A
C 43	231.5	6.4	375	7	US-11-096-568A-23618	Sequence 23618, A
44	231.5	6.4	526	7	US-11-096-568A-16997	Sequence 16997, A
45	230.5	6.4	702	7	US-11-096-568A-14646	Sequence 14646, A

ALIGNMENTS

RESULT 1
US-11-096-568A-21828
; Sequence 21828, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 21828
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(386)
; OTHER INFORMATION: Ceres Seq. ID no. 12406424
US-11-096-568A-21828

Alignment Scores:
Pred. No.: 1.69e-10 Length: 386
Score: 294.50 Matches: 127
Percent Similarity: 37.4% Conservative: 26
Best Local Similarity: 31.1% Mismatches: 117
Query Match: 8.1% Indels: 139
DB: 7 Gaps: 28
US-10-620-914-44 (1-1947) x US-11-096-568A-21828 (1-386)

QY 1684 TGGTAGGTGGCGGCTTTGAGCTCTCCATGAAGAAGTTGGTGG-----AGA 1637

```
Db 35 TrpArgArgCysAlaAlaAlaThrArgProGlnArgArgHisArgTrpProAspArgArgArg 54
QY 1636 CGG-----TCAGGTTGTCCACACCGCCTCTTTGAGGGTGGGAAGCGCGCTCGCGCA 1583
Db 55 ArgProGlyProGlyArgAlaProArgGlyAlaAlaGlyProArgArgPro-----71
QY 1582 GGTAGGTGGGCGCATTGTCCG-----1562
Db 72 ---GlyGlyGlySerThrArgArgLeuLeuAlaAlaArgGlnProGlnMetProArg 90
QY 1561 ---CCAGGAACCTCCCGTGGAGGAGTGTAGTAGAAGTAGTCTGTGTCGGCAGCGCG 1505
Db 91 GlySerGlyAlaSerArgProGlySerCysSer-----SerArgGly 104
QY 1504 AGTTCTCGCCGACCGCTCATGTCGGCGGATGTAGTCTCAATGGGATCGCTCG 1445
Db 105 SerThrGlyProArgArg---TrpProAla---AlaGlyGlyArgTrp-----118
QY 1444 CCTGATCAGCGCTACTGTTCGCGCA-----CGCGCGCGCGAACC 1400
Db 119 -----ArgThrProCysProCysAlaSerProSerProArgAlaArgPro 134
QY 1399 ACAGCAGCGCTTGTGAAGAGCACCGGCTCAGCACTTCAGCAAGCAGCACACCGAGCG 1340
Db 135 ThrAlaAlaGlyGlyAlaThrSerThrGlyCys-----SerThrAlaAlaAla 151
QY 1339 ---GTTTGGGCGGCTTCTTCCAGAGTGGATGAGCATGTTGTCGTCACAGCGCGCT 1283
Db 152 ArgThrTrpProArgArgAlaGlyThrGlyCysThrThrCysSerPro-----167
QY 1282 GCTCTCCATTGTGGCGCGTGTGGCGGCGCTTGACGGCTTGGCCAGTTCGCCAGTCCACCA 1223
Db 168 -----ArgSerSerProLeuProSerPro 175
QY 1222 -----CGSCCAGCG-----1190
Db 176 ProThrAlaProArgArgProGlyProProArgTrpAsnAlaAlaSerArgGlyTrpThr 195
QY 1189 CCA-----TGCGCGCTGTAGTACAGCGCGTGTGGAAGTACAGA-----1148
Db 196 ProArgArgCysArgProAlaGlyAlaProProAlaAlaAlaThrArgThrSerAlaThr 215
QY 1147 -----GGCGCTGGACAGCAAGTGTGGCTGTTGCGTTCGCGCA 1112
Db 216 ThrTrpAlaProProTrpTrpProSerTrpArgSerAlaAlaSerTrpTrpProThr 235
QY 1111 GGAAGGCGCGAGCTTCTCTCGTACAGCTCTCAATGC-----GCGGCTGACGCGCT 1058
Db 236 -----ValAlaThrAlaAlaArgCysAlaAlaGlyAlaAlaArg 249
QY 1057 GCSCGAACAGCTGCCACAGCTCTCAACTCCAGCTGCT-----GAATGG-----1013
Db 250 ArgArgCysArgCysArgArgThrThrSerProThrAlaProThrSerTrpSerGlySer 269
QY 1012 ---CCACCTTCTTACAGAGCGGCGCTGCGGGTGGAGTTCAGTCCACCGACCA 956
Db 270 ArgProArgAlaAlaAlaSerSerGlyArgAlaAlaArgAlaCysTrpAlaSerSerPro 289
QY 955 CTGGCGCGCGCTGCGACAGCAGTTCAGGCGATTGCGAGCGCGCTAGTCAGGGTCA 896
Db 290 -----CysProAlaProSerGlyThrAlaThrSerSerArgThr-----Ser 303
QY 895 GCACCGTGTCTTGGGGTTGATCTCCATCACCTCCATATCCG-----GCTCGGGTCTCT 842
Db 304 ArgProCysProArgSerArgSerProThrAlaProThrProThrSerAlaSerSerSer 323
QY 841 CCCACGACTGCTGTACAGGAAGTGGCGGGAACATGGCGCGCGCTCCACGCGTCTCT 782
Db 324 ProAlaThrAla-----SerGlyThrTrpLeuAlaThr-----ArgPro 336
QY 781 CTGTGACGGCGTGGCCACAGCTGGGCGAGCGCGCAATCCACAGCTAGTAGGGGCGCGCA 722
```

```
Db 337 ArgAla-----ArgTrp-----Arg 341
QY 721 GCCACGCGCAGTAGGCGATCGAACCTCTGG-----TGTTCTGCT 683
Db 342 AlaProAlaSerAlaGlyGlyProGlyGlyAlaProArgProArgProArgCysSerPro 361
QY 682 CCCACACGCGCTCCAGCTTCTGTCTCCA 656
Db 362 SerTrpProSerProGlyAlaAlaPro 370

RESULT 2
US-11-096-568A-21828
; Sequence 21828, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 21828
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(386)
; OTHER INFORMATION: Ceres Seq. ID no. 12406424
US-11-096-568A-21828

Alignment Scores:
Pred. No.: 1,32e-09 Length: 386
Score: 279.50 Matches: 126
Percent Similarity: 36.4% Conservative: 31
Best Local Similarity: 29.2% Mismatches: 123
Query Match: 7.7% Indels: 151
DB: Gaps: 22

US-10-620-914-44 (1-1947) x US-11-096-568A-21828 (1-386)
QY 711 CGTGGCGTGGTGGCGCGCCCTACTAGTGTGAGTGGCGCGCTGCCAGCGTTGGCA 770
Db 37 ArgCysAlaAlaThrArgProGlnArgArgHisArgTrpProAspArgArgArgPro 56
QY 771 CGCCCTGCACAGAGCGCGTGGAGCGCGCCCGCCATGTTCCGCGCCACCTTCTCTACAC 830
Db 57 GlyProGlyArgAlaProArgGlyAlaAla-----66
QY 831 GCAGTCTGGGAGGACCGCGAGCGGATATGAGGTGATGAGATCAACCCACAGACAC 890
Db 67 -----GlyProArgProGlyGlySerThr-----76
QY 891 GGTGTGACCTGACTAGCGCGCGGTG-----CAATGCCCTGAA 929
Db 77 -----ArgArgLeuLeuAlaAlaArgGlnProGlnMetProArg 90
QY 930 CTGTGTGTGAGGGGGCGCGCGAGTGTGTCGTGCTGCAACCCCGCGCAGTCCGC 989
Db 91 GlySerGlyAla---SerArgProGlySerCysSer-----SerArgGly 104
QY 990 GCTTCTGAGCTGAAGAAGTGGCCATTCAGCAGCTGGAGTTTGAGGAGCTGTGGCAGCT 1049
Db 105 SerThrGlyPro---ArgArgTrpProAlaAlaGlyArgTrpArgTrpProCysProC 124
QY 1050 GTTCCGCGAGGGCGTGCACCGCGCATTCAGGAGCTGTACGAGAAGAGCTGGCGCCTT 1109
Db 124 ysAlaSer-----ProS 128
QY 1110 CTGTGCGGAACACCGACCACTTCTGTGTCGAAGCGCTCTGTACTTCCAGCAGCGCT 1169
Db 128 erProArgAlaArgProThr-----AlaAlaGlyGlyAlaThrSerThrGlyC 145
```


Db 1234 ThrThrThrAlaCysThrThr-----CysThrAlaCysThrThrThr 1247
Qy 1217 TGGCGTGGTCTGGGACTGGGCAAGACCGTCAAGCGCTCGCCACACGCGCCCAATGG 1276
Db 1247 ----- 1247
Qy 1277 AGGAGACGCGCTGTGGGACAGCAACATGCTCATCCACTTCGTGAAGACGGGCCCA 1336
Db 1248 ---AlaAlaThrCys---AlaAlaThrCysAlaThrThrAlaAlaAlaGlyGlyAla 1265
Qy 1337 AGCGCTGGTGGCTGTGCTCAAGTTCGTGAGCC----- 1372
Db 1266 AlaThrCysCysAlaAlaAlaAlaAlaGlyCysAlaAlaAlaCysAlaGlyAlaGlyThr 1285
Qy 1373 TGGTGTCTTCAACAAGCGCTGTGTGTTGCGGCGCGCTGCGGCAAGCAGTACG 1432
Db 1286 ThrCysAlaGlyThrCysThrCysAlaThrThrGlyCysAlaAlaGlyGlyThr 1305
Qy 1433 CGCTGATCAAGCGGAGCGATCCCATTTGAACTACATCG----- 1474
Db 1306 ---AlaAlaAlaThrAlaThrCysAlaThrThrAlaAlaThrThrGlyGlyAla 1323
Qy 1475 -----CGCGCACCATGACGCGTGGCGGAGAACTCGCAGTGGCGAAGCAGAACTACT 1528
Db 1324 AlaGlyThrAlaGlyThrThrAlaAlaAlaThrGlyThrCysThrCysAlaThrThr 1343
Qy 1529 TCTACTCAACTGCTCACCAGCAAGTTCTTCGCGCAGCAACTGCGCCACCTACCTGGCGG 1588
Db 1344 GlyThrThrThrThrAlaThrThrGlyAlaCysAla-----CysAla 1357
Qy 1589 AGCGGCGCTCGCACCCCTCAAGAGTGGCGGTGGGAGCAACTGACCTGCTCCACCACT 1648
Db 1358 ThrCysThrAlaThrAlaThrAlaThrAlaCysAlaThrThr----- 1371
Qy 1649 TCTTCATGGAGGCTCAAGCGCGCACCTACACCAAGGTGATCTGATGACACACGTTGG 1708
Db 1372 -----ThrGlyThr-Gl 1375
Qy 1709 ACTGCTGATATGCGGTGGCCACAGAGTGGCGAGTGGCGGCTGGCCAGAGTGGCG 1768
Db 1375 YAlaAlaGlyCysAlaAlaGlyAlaAlaAlaGlyAlaAlaAlaAlaAlaAlaAla 1395
Qy 1769 GCGGCGGATGCTCATCTGCGCTCGGCTCGCTCCCTCAGCGCGCTACGCGGAGTATCC 1828
Db 1395 agly--CysThrThrCysGlyThrAlaThrGlyCysCysAlaThrThr----- 1410
Qy 1829 AGAGCGGCTTCGACGTGGCTGCATCGCGCGCCACTCAGGCTACATGACCGCG 1888
Db 1411 ---AlaAlaThrThrAlaAlaCysAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 1428
Qy 1889 TCAACATGT-----ACAGCTCTTCT 1909
Db 1429 ThrThrCysAlaGlyThrAlaCysThrGlyAlaThrThrGlyCysAlaThrAlaCysAla 1448
Qy 1910 ACATGCGCGCGGAGGCGCCAGAGGCAACT 1945
Db 1449 AlaGlyAlaThrGlyCysAlaThrGlyThrThrThr 1460

RESULT 4
US-11-182-016-31
; Sequence 31, Application US/11182016
; Publication NO. US20060019294A1
; GENERAL INFORMATION:
; APPLICANT: SUGEN, INC.
; TITLE OF INVENTION: TYROSINE KINASE SUBSTRATE (TKS) PROTEINS
; FILE REFERENCE: 038602/0102
; CURRENT APPLICATION NUMBER: US/11/182,016
; CURRENT FILING DATE: 2005-07-15
; PRIOR APPLICATION NUMBER: US/09/958,359
; PRIOR FILING DATE: 2002-02-05
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 31

; LENGTH: 495
; TYPE: PRT
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Tks 118
US-11-182-016-31
Alignment Scores:
Pred. No.: 4.95e-09 Length: 495
Score: 270.00 Matches: 150
Percent Similarity: 30.7% Conservative: 40
Best Local Similarity: 24.3% Mismatches: 171
Query Match: 7.5% Indels: 257
Db: 34 Gaps: 34

US-10-620-914-44 (1-1947) x US-11-182-016-31 (1-495)
Qy 1811 GCGCGCTGAGGAGCGGAGCGCGCAGATGATGCGCGCGCGCAACTGTTG--- 1755
Db 2 GlyGlyCysArg-----LysAspValTrpGluSerProGlyHisCysCysIleTrp 18
Qy 1754 -----GCCAGCCTCGCGCAGCTCGTTGGCCACGGC----- 1722
Db 19 GlyAspArgLeuSerLeuProSerArgGluProGlyLeuIleLeuThrGlyGlyPro 38
Qy 1721 -----ATATCCAGCCAGTCCACGTGGTCCATCAGAAATCACCTTG 1683
Db 39 IleProGlyArgValLeuAsnValAsnSerGlnAla----- 50
Qy 1682 GTGTAGTGGCGCTTTTGGAGCTCTCCATGAGAAAGTTGGTGGAGAGCGTCAGTTGTTC 1623
Db 51 -----SerSerSerSerAsnLysGluAlaMetSerGluArgGlyArg 64
Qy 1622 ACCAGCCACTCTTGAAGGTGGCGAAGCCCTCGCGCAGGTAGTGGGCGCAGTTGTCG 1563
Db 65 AlaArgCysAlaLeuSerLeuThrGlnAla--ProArgSer----- 78
Qy 1562 CGCAGGAAGTTCGCGGTGAGGACAGTTGTAGTAGAAGTAGTTCTGCTTCGCGACGTCGAG 1503
Db 79 -GlnGlyThrCysGlnAsnGlyHisProAlaHisSerHisAlaThrSerLeuArgArgSe 98
Qy 1502 TTCTCCGCGCAGCGCTCCATGCTGCGCGCATGATGATTTCTCAATGGGATGCCGTCGCC 1443
Db 98 r-----ProArg----- 100
Qy 1442 TTGATCAGCGCTACTGCTTCGCGCGCACGCGC-----CGCCGAACACACAGCAGCGC 1389
Db 101 -----CysProGlySerArgGlyGlnArgArgSerLeu----- 111
Qy 1388 TTGTTGAAGACACACAGGCTCAGCACTTGACGAACACAGCCACACAGCGGCTTGGGCGCG 1329
Db 112 -CysArgArgLeuProGlySerArgThrGlyHisGlyHisArgAlaLeuAlaHisAlaPr 131
Qy 1328 TTCTTCAGAAAGTGAATGATGATGTTGTCGCCACAGACGCGCTGCTCCTCCATTGTG 1269
Db 131 oglyProGluCysGlyGlyGlnCysAlaGlnSerGlnHisLeuAlaAlaProValGl 151
Qy 1268 GCGCGTGGCGAGCGCTTGACGGTCT-----ArgSerProAlaLeuGluValProHisThrGlnProPr 168
Db 151 yAlaAlaArgArg----- 1240
Qy 1240 -----TGCCAGTCCGACCCACCGC----- 1220
Db 168 oLeuLeuProAlaProAspMetGluGluCysTrpAlaProAlaProGlnGlnGlyAspLe 188
Qy 1219 -CCAGGCACCTGCAGCACCCAGCAGCTTGCCCATGCGCGCTGTAGT---ACAGGCGC 1164
Db 198 uProTrpAlaLeuArgTrpGlnGlnLeuAlaLeuCysGlnSerGlySerLeuValGlyAl 208
Qy 1163 TGCTGAAGTACACAGCGCT-----TG 1140
Db 208 aAlaSerAlaGlyArgGlySerSerAlaCysGlnAlaGluAspTrpArgTrpArgTrpTr 228

```

QY 1139 GACGAGAGTTGTGGCTGG-----TTTGGCAGAGGAGGCGGCGAGCTTCTTCTGTGACAGC 1083
Db      |||      |||      |||      |||      |||      |||      |||      |||      |||
228 pThr-----TTPProSerProSerAlaLysSerProSerGlyAlaArgGlnTh 244
QY 1082 TCCT-----CAATCGCGGG 1068
Db      |||      |||      |||      |||      |||      |||      |||      |||      |||
244 rProAlaHisPheGlySerLeuLysProLeuLeuAlaAlaLapProThrSerCysSerG 264
QY 1067 TGCAGCGCTCGCGAAGAGCTGCCACAGCTCCTCAACTCCAGCTGCTGAATGCCACAC 1008
Db      |||      |||      |||      |||      |||      |||      |||      |||      |||
264 yGlyTTPProArgHisProAla-----ProAspSerAlaAlaProGlyValGlnPr 281
QY 1007 T-----TCTTCAGTCCAGAGCGCGGAGCTGCCGCGGGTTGCAGTCCACGACACACC 954
Db      |||      |||      |||      |||      |||      |||      |||      |||      |||
281 oHisAlaAlaHisAlaProGlyAlaProAlaProAlaAlaValProProArgProAl 301
QY 953 TGGCGCGCGCCCTGCA----- 938
Db      |||      |||      |||      |||      |||      |||      |||      |||      |||
301 aAlaPheProProAlaAspGlySerProProSerLeuLeuTTPProArgSerCysLe 321
QY 937 -----CCAGCAGGTTTC 927
Db      |||      |||      |||      |||      |||      |||      |||      |||      |||
321 uTyR-GlnProPheSerGlnThrArgHisTTPSerSerGlyThrGlnSerProLeuGlyPr 341
QY 926 AGGCGATTGCCAGCGCGCTAGTCAGGTCAGCACCGTGT----- 887
Db      |||      |||      |||      |||      |||      |||      |||      |||      |||
341 oGly---ValProArgProGlySerGlyHisSerProCysGluSerCysSerTrpHisLe 360
QY 886 -----CCTTGGGGTTGATCT-----CCATCACTCCATATCGGCTCGG-- 848
Db      |||      |||      |||      |||      |||      |||      |||      |||      |||
360 uLysProTTPProProSerProCysThrGlnAlaProHisProProArgProValArgTr 380
QY 847 -----GGTCCTCCACAGCTCGGTGTACAGGAAGGTGGCGCGGAAACATGGCGCGCGC 795
Db      |||      |||      |||      |||      |||      |||      |||      |||      |||
380 pSerHisGlyProProSerGlySerTTPPro-----TTPCysArgGlyTTPHisArgLe 398
QY 794 TCCAGCGGCT-----CCTGTGAGGCGGTGGCGCAACGCTGGGCGCGCGCCA 747
Db      |||      |||      |||      |||      |||      |||      |||      |||      |||
398 uProSerAlaHisArgSerArgProArgLeuSerSerGlyGlnIleTTPAla----- 415
QY 746 ATCCACACGTAGTAGGGGCGCGCAGCCAGCGTAGGGATCGAACCTGGGTGTTTC 687
Db      |||      |||      |||      |||      |||      |||      |||      |||      |||
416 -----ValGlnSerTTPGly---ProSerLeuCysArgArgThrSerProSerArgCy 433
QY 686 TGCTCCACACGCGCTCCAGCTTCTGTCTCAGGTAGGCGGCGCTCGGCGCGCGATGCA 627
Db      |||      |||      |||      |||      |||      |||      |||      |||      |||
433 sAlaProProSerProProGlyHisProProLeuCysGlnProArgGly---CysHi 452
QY 626 ATGTTGTGATGTCGAAGATCGATCGCCAGAGAAACGGCGCGACCGGCGATCTGGCGC 567
Db      |||      |||      |||      |||      |||      |||      |||      |||      |||
452 sCysCys----- 454
QY 566 AGGGCAGGTGCTACTTTCGCTCAGCTAGAGTTCGGCAAGCCACCGCGCTTGTGG 507
Db      |||      |||      |||      |||      |||      |||      |||      |||      |||
455 -----CysLeuHisArgArgGluProSerArg----- 463
QY 506 GACAGGTACGAGCAAGCTGTGTCGATCAGCTGTGTGGAACGCTGGAATCATCGTGAGCGAG 447
Db      |||      |||      |||      |||      |||      |||      |||      |||      |||
464 -SerGlyThr-SerArgPro----- 469
QY 446 TAGGAGAAGGTGATGAGCGTTCGGGTGCCCTCAGGGGGCGCAATGCGAAGCGTTCGCC 387
Db      |||      |||      |||      |||      |||      |||      |||      |||      |||
470 -----ProAlaArgArgPr 474
QY 386 T-----CCACGAGCTGACATTTCTTCAGCGCTTGGCTTCGGCT 347
Db      |||      |||      |||      |||      |||      |||      |||      |||      |||
474 oLeuAlaAlaLeuAlaArgSerGlySerGlySerProProTTPProAlaPro 491

```

RESULT 5

US-11-096-568A-19503

; Sequence 19503, Application US/11096568A

; Publication No. US2006004820A1

; GENERAL INFORMATION:

```

; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; TITLE OF INVENTION: Thereby
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096.568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 19503
; LENGTH: 428
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(428)
; OTHER INFORMATION: Ceres Seq. ID no. 12373378
US-11-096-568A-19503

```

Alignment Scores:

```

Pred. No.: 6.87e-09 Length: 428
Score: 267.50 Matches: 126
Percent Similarity: 41.6% Conservative: 30
Best Local Similarity: 33.6% Mismatches: 103
Query Match: 7.4% Indels: 116
DB: 7 Gaps: 30

```

US-10-620-914-44 (1-1947) x US-11-096-568A-19503 (1-428)

```

QY 205 CGCTCGAAGCTCATCTCTGGTTGACCTGGTGGT---GGCACTGGGAGAAATGTCGATATG 261
Db      |||      |||      |||      |||      |||      |||      |||      |||      |||
131 ArgGlyAsnIleLeuSerTyrAsnLeuGlyGlyHisGly-LeuGlnArg----- 146
QY 262 ATGGCTGATTACATCGACCTGGCGAGTTCAGATCCATCTACGTGG---TCGACCTGTGC 318
Db      |||      |||      |||      |||      |||      |||      |||      |||      |||
147 -----ArgAlaSerSerProSerIleTTPProAlaThrCysSe 159
QY 319 CACTCGCTGTGCGAGGTGGCCAAAGAGAGCGGAGCCAGGCTGGAGAATGTCCAG 378
Db      |||      |||      |||      |||      |||      |||      |||      |||      |||
159 r-----ArgProAlaAla-----Ar 164
QY 379 GTGCTGGAGCGCAGCGCTTGCCAAATTTGCGCCCTCGAGGCGACCGCGCTCATCACC 438
Db      |||      |||      |||      |||      |||      |||      |||      |||      |||
164 gAlaTTP---ProTTPTrpSerAlaArgArgPro-----SerProSerArgGlyThrPr 181
QY 439 T-----TCTCCTACTCGCTCAGCATGATTCCACCGTTCCACAAACGTCATCGACCG 489
Db      |||      |||      |||      |||      |||      |||      |||      |||      |||
181 oGlySerGlyAlaProCysSerSerArg-----ThrProSerSerGlyPr 196
QY 490 GCTTGCTCTGACCTGT-----CCCAAGACGGC 516
Db      |||      |||      |||      |||      |||      |||      |||      |||      |||
196 oAlaAlaAlaArgCysCysProThrGlyAlaAlaThrSerThrAlaProSerThrSe 216
QY 517 CTGG-----TGGCGGTGGCGGCTTCTACGTGAGGCGCAAGTACGACCTGCCCTCGGC 570
Db      |||      |||      |||      |||      |||      |||      |||      |||      |||
216 rTTPSerThrTTPCysAlaArgThr-----ArgAlaArgThrThrAlaProSerAl 233
QY 571 CAGATGCTCTGTGTGGCGCGCTTCTTCTGTGGCGATCGATCTTCGACATCGACAAATGAC 630
Db      |||      |||      |||      |||      |||      |||      |||      |||      |||
233 aProCysThrArgArgThrSerSerGlySerArg----- 245
QY 631 ATCGGCGCGAGCGCGCGCTTACCTGGAGCAGAGCTGAGGCGCGTGTGGGAGCAGAAC 690
Db      |||      |||      |||      |||      |||      |||      |||      |||      |||
246 -AlaCysProSerAlaAlaThrCysTTP-----ArgTTPAlaAlaThrArgSerArgPr 263
QY 691 ACCAGGGTTCGA---TCCCTACTCGTGGCTGGCTGGCGCCCTACTACGTGTGGATT 747
Db      |||      |||      |||      |||      |||      |||      |||      |||      |||
263 oThrSerProArgTTPGlyProTTPCys-----CysProSerProSerCysSerSe 281
QY 748 GGCC---GCCTGCCAGCGTTGGCCACGCCCTCGACGAGGAGCGGTGGAGCGCGCCGCC 804
Db      |||      |||      |||      |||      |||      |||      |||      |||      |||
281 rSerArgAlaCys-----Cys-----SerAlaThrCysSerArgPr 293
QY 805 ATGTTCCCGCCACCTTCTCTGTATACGCACTGTGGGAGAGCCCGGACCGGATATGGAG 864
Db      |||      |||      |||      |||      |||      |||      |||      |||      |||

```



```

Db 744 SerSerSerTrpGln---ArgAlaTrpGlnArgProHisSerArgAlaAlaGlyLeu 762
Qy 827 ACAGCAGTCTGGGAGGACCCCGGCGGATA---TGG----- 862
Db 763 ProArgSerArgAlaArg-----AlaArgIleSerTrpLysValProAlaLeuGlySer 780
Qy 863 -----AGTGTAGAGATCAACCCAGGACACGGTGTGCTGACCTGACTAGCG 910
Db 781 GlyAlaArgArgAlaSerAlaAlaSerThrAlaSerArgProGlyProAlaProAla 800
Qy 911 GCGGCTCAAGCCCTGGAAC-----TGCTGG-----TGCAAGGGG 946
Db 801 SerArgAlaLysProSerSerThrHisSerCysAlaCysTrpLysLysArgCysSerArg 820
Qy 947 CCGGCCAGGTGG-----TGTCGGTGGACTGCAACCCCGCAGTGGCGGC 991
Db 821 ArgAlaArgTrpArgProLeuSerThrCysSerSerArgCysSerLysLysSerThr 840
Qy 992 TTCTGG-----AGTCAAGAAGGTGGCCATTCCAGCAGC 1024
Db 841 PheTrpGluValAlaPheSerSerArgAlaArgAlaSerArgAlaTrpAlaArgSer 860
Qy 1025 TGGAGTTTGGAGGACGTGGGAGCTGTTCCGCGAGGGCGTGCACCGCGATTGAGGAGC 1084
Db 861 Trp-----ThrSerAlaLeuLysAlaArgAsnArgSerValSer 874
Qy 1085 TGTACGAGAAGAGCTGGCGCTCTCTGTCGCAAAACCCAGCCCAACTTCTGTGCCAAGC 1144
Db 875 CysArgAlaAlaPheLysValSerProGlyIleAlaLysLeuAlaSerCysSerProGly 894
Qy 1145 GCCTCTGTACTCCAGCAGCGCTGTACTACCAAGGCGGCGATGGCAAGCTGTGCTGGG 1204
Db 895 ProGluSerHisCysAsnThr---CysTrpSerArgArgHisCysArgCysSerCys 913
Qy 1205 TGCTGCAAGTCTGGCCGTGTGTC-----TGGGACTGGCGCAAGACCTCA 1249
Db 914 CysCysThr-----CysArgLeuGluArgArgThrSerTrpMetAlaSerSer 929
Qy 1250 AGCCCTCGCCCAACCGCGCCCAACATAGGAGGAGCAGCGCTCTGTGGGACA----- 1300
Db 930 ThrSerTrpSerValAlaGlyProTrpProSerSerLeuLeuLeuGlyValGluArgSer 949
Qy 1301 -----GCAACATGCTCATCTCACTTCGTGAAGAACGGGCCCCAAGC 1339
Db 950 LeuIleArgMetAlaProProSerLeuCysSerAlaValSer---ArgGlySerAlaSer 968
Qy 1340 CGCTGGGTGCG-----TGTTGCTCAAGTTCGTGAGCTGTGTGCTCT 1381
Db 969 ThrPheCysSerGlyLeuGlyIleProThrAlaSerSerSer----- 983
Qy 1382 TCACAAGGCGCTGCTGTGGTTCGCGCGCGCGTGCAGGCAAGCAGTACGCGCTGATCA 1441
Db 984 -----GlySerAlaAlaProSerSerSerSerMetAla----- 994
Qy 1442 AGCGGAGCGGATCCCATTTGAGAACTATACATCGCGCGCACCATG----- 1486
Db 995 -----CysArgThrGluProSerThrProTrpHisArgArgHisSer 1008
Qy 1487 -----ACGGCGTGGCGGAGNACTCGCAGCTGCGCAGCAGAACTACTTCTACTACAACT 1540
Db 1009 LeuThrThrSerTrp-----CysIleSerThrGlnLeuGlyGluGlySer 1023
Qy 1541 GCCTCAGCGGCAAGTCTCTGCGGCAAACTGCCCCCACTACCTCGCGGCGCGCTTCG 1600
Db 1024 LeuProProAsnSerHisCysSerSerSerGlyLeuAlaThrLeuPheArgSerPheSer 1043
Qy 1601 CCACCTCAAGAGTGGCGTGGTGGCAACCTGACCGCTCCACCAACTTCTTCATGAGG 1660
Db 1044 AspSer-----ThrSerAlaProAsnProGlnSerSerValGlyArg 1057
Qy 1661 AGCTCAAGCGCGCACTACACCAAGGTGATTCTGATGGACCAGCTGGAGTGGCTGATA 1720

```

```

Db 1058 SerSerMetArgIleSerSer-----ArgThrArgGlyGly----- 1069
Qy 1721 TGCCCGTGGCCAAAGAGTGGCCGAGTGCCTCGGCAAGCAGGTTTCGCGCGCGGCATCG 1780
Db 1070 -----SerProGluSerSerTrpLeuLysAlaGlyMetSerAlaGlyGly 1084
Qy 1781 TCATCTGGCGTCCGCTCCCTCAGCCCGCTCAGCCGCTACGCGGAGCTGATCCAGAGGGGCT 1840
Db 1085 SerGlyGlyAlaArgArgSerSerArgThrAspSerProSerValCysArgSerGlyLeu 1104
Qy 1841 TCGACG 1846
Db 1105 SerSer 1106

RESULT 11
US-11-096-568A-20771
; Sequence 20771, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 20771
; LENGTH: 413
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(413)
; OTHER INFORMATION: Ceres Seq. ID no. 12387083
US-11-096-568A-20771

Alignment Scores:
Pred. No.: 1,46e-08 Length: 413
Score: 262.00 Matches: 130
Percent Similarity: 38.7% Conservative: 42
Best Local Similarity: 29.3% Mismatches: 138
Query Match: 7.2% Indels: 134
Gaps: 26
DB:

US-10-620-914-44 (1-1947) x US-11-096-568A-20771 (1-413)
Qy 284 CGAAGTTCAGTCCATCTAGCTGTGTCGACCTGTGCCACTCGCTGCGAGGTGG----- 337
Db 72 ArgGlnAlaAlaAsnGlnProTrpSerAlaSerArgSerCysSerSerTrpAlaSer 91
Qy 338 -----CCAAGAAGAGCGCAAGGCCAAGGCTGGAAGATGTCAGG 379
Db 92 SerAlaSerThrAlaSerProArgArgSerSerGluAlaArgThrAlaAlaSerArg 111
Qy 380 TCGTGGAGGCGGAGCTGGCCAAATTTGCGCCCTGAGGCGACCGCA-----CGC 430
Db 112 AlaSerArgPro-----ProSerArgCysProSerProArgPheSer 125
Qy 431 TCATCAGCTTCTCTACTGCTCAGCATGATTCACCGTTTCCACAAGCTCATCCAGCAGG 490
Db 126 SerCysThrSerProSerArgThr-----ArgSerSerProThrSerThr--- 140
Qy 491 CTGTGCTGACTGTCTCCCAAGAGCGCTGTGGCGGTGCGCACTTCTACGTGAGCGGCA 550
Db 141 -----ProThrThrArgProTrpThrHisProProArgAlaProThrArg 155
Qy 551 AGTACGACCTGCGCTGCGCGCAGATGCCCTGTGTCGCGCCGCTTCTTCTGCGGATCGATCT 610
Db 156 SerAlaAlaCys-----CysPro-----Ala 162
Qy 611 TCGACATCGCAACATTTGACATCGGCCCCCGCGCGCTACCTCGAGCAGAGCTGG 670
Db 163 SerProProThrGlySerProCysSerSerSerProProThrThrSerArgSerSer 182

```

		FILE REFERENCE: 68175	
		CURRENT APPLICATION NUMBER: US/10/894,592	
		CURRENT FILING DATE: 2004-07-20	
		NUMBER OF SEQ ID NOS: 3	
		SOFTWARE: PatentIn version 3.2	
		SEQ ID NO 3	
		LENGTH: 1356	
		TYPE: PRT	
		ORGANISM: Homo sapiens	
		US-10-894-592-3	
		Alignment Scores:	
		Pred. No.:	1,87e-08
		Score:	261.00
		Percent Similarity:	33.6%
		Best Local Similarity:	24.3%
		Query Match:	7.2%
		DB:	6
		US-10-620-914-44 (1-1947) x US-10-894-592-3 (1-1356)	
Qy	473	ACAACTGTCATCGACCGCTTCTGCTACCTGTCCTCCCAAGACGGCTGGTGGCGTTGCGG	532
Db	54	ThrAlaAlaCysThrThrCysAlaThrThrCys	64
Qy	533	ACTTCTACGTGAGCGGCAAGTACGACCTGCCCTGCGCCAGATGCCCTGTGTCGCGCCGTT	592
Db	65	---ThrThrCysAlaGlyThrAlaCysAlaGlyAla---CysAlaGlyThrGlyThr	82
Qy	593	TCTTCTGGCGATCGATCTTCGACATCG	640
Db	83	ThrThrGlyThrGlyThrGlyThrThrCysCysThrGlyThrGlyThrGlyCys	102
Qy	641	AGCCCGCGCTACCTGGAGCAGAGCTGGAGCGG	697
Db	103	ThrAlaAlaCysThrThrThrAlaAlaGlyAlaAlaGlyCysGlyGly	120
Qy	698	GTTCGATCCCTACGTGCGGCTGCGGCCCTCTACTAG	739
Db	121	ThrThrThrCysThrCysAlaGlyAlaAla	138
Qy	740	---TGTGATTTGGCGCTGCCAGCGTTGGCCAGCCCTGCACGAGGAGCGCGTGG	793
Db	139	AlaCysCysAlaThrGlyGlyCysAlaCysAlaAlaAlaCysThrGlyThrGlyAla	158
Qy	794	AGCGCGCCCATCTTCCCGCCACCTTCTGTCACGACGTCGTCGGGAGGACCCCGAGC	853
Db	159	CysGly	171
Qy	854	CGGATATGGAGGTGATGGAGATCAACCCCAAGGACACGCTGCTGACCTGACTAGCGCG	913
Db	172	---AlaGly	173
Qy	914	GCTGCAATGCCCTGCAACCTGCTGTGTGAGGGGCGCCAGGTGTGTCGGTGGACTGCA	973
Db	174	ThrAla	189
Qy	974	ACCCGCGCAGTCGCGCTTCTGGAGCTGAAGAGGTGCCATTTCAGCAGCTGGAGTTG	1033
Db	189	---	189
Qy	1034	AGGAGCTGT	1054
Db	190	---ThrCysCysThrAlaThrThrGlyCysAlaCysAlaGlyAlaAlaAlaThrCysAla	208
Qy	1055	GCGAGGCGTGCACCCGCGCATTTGAGGAGCTGTAGAGAGAGCTGGCGCCCTTCCTCT	1114
Db	209	GlyAlaGlyCysCysThrCysAlaThrThrCys	220
Qy	1115	CGCAACACGACCACTTCTGTCACAGCGCTCTGTTACTTTCAGACAG	1165
Db	221	CysThrThrAlaGlyThrAlaGlyThrGlyAlaAlaGlyCysAlaThrThrGlyAlaThr	240

RESULT 12
US-10-894-592-3
; Sequence 3, Application US/10894592
; Publication No. US20060019335A1
; GENERAL INFORMATION:
; APPLICANT: Du, Chunying
; APPLICANT: Yang, Qi-Hang
; TITLE OF INVENTION: Compositions and Methods to Promote c-IAP Autodegradation

QY 1166 -----GCTGTACTACAGGCGGCA 1186
Db 241 GlyAlaGlyGlyAlaGlyAlaGlyCysAlaGlyThrGlyThrCysThrThrThrGlyGly 260
QY 1187 TGGCGAAGCTGTCTGGG-----TGCTGAGTGGCTGGCGGTGGC 1228
Db 261 ThrAlaAlaCysAlaGlyAlaThrAlaGlyCysAlaCysCysThrCysThrAlaGlyCys 280
QY 1229 TGGGACTGGCAAGACCGTCAAGCGCTCGCCACAGCGGCCCAATGAGGAGGAGCGGCC 1288
Db 281 -----ThrThrThrCysThrCysThrCysThrCysAlaGlyAlaCysCys 295
QY 1289 GTCTGTGGGAGCAACA---TGCTCATCCACTTGTGAGNACGGGCCCAAGCGCTGG 1345
Db 296 AlaCysAlaThrAlaThrGlyCysGlyThrGlyAlaThrGlyAlaAla----- 313
QY 1346 TGTGCTGTCTCAAGTCTGTGAGCTGTCTCTTCAACAGCGCTGTGTGTGTCG 1405
Db 314 ---GlyCysThrAlaThrThrAlaCysThrGlyAlaAlaThrAlaThrAlaCysThrAla 332
QY 1406 GCGCGCGGTGCGGCGCAAGCAGTACGCGTGTATCAAGCGCGGCGCATCCCATTTGAGA 1465
Db 333 AlaGlyGlyCysThrGlyThrThrAlaThrAlaCysCysThrThrAlaAlaCysThr 352
QY 1466 ACT-----ACATCGCGCGCACCATGGACGGCTGGCGGAGNACTCGCAGC---TGC 1513
Db 353 ThrCysThrCysThrThrThrAlaCysCysGlyAlaCysAlaAlaThrAlaThrAlaCys 372
QY 1514 CAAGCAGAACTTACTTACTACTACACTGCTCAGCGCAAGTCTCTGCGGACAACTGCC 1573
Db 373 AlaAlaGlyThrThrAlaCysThrThrGlyGlyGlyAlaAlaAlaAlaThrGlyAla 392
QY 1574 CCACCTACTGCGGAGCGGCTTCCGCCACCTCAAGAGTGGCGTGGGACCACTGA 1633
Db 393 AlaThrThrCysAla-GlyAla----- 399
QY 1634 CCGTCTCACCAACTTCTTCTGAGGAGCTCAAGCGCGCACCTACACCAAGGTATC 1693
Db 400 -----GlyGlyAla----- 405
QY 1694 TGATGACCACTGGACTGGCTGGATATGCCGCTGGCCCAAGAGTGGC----- 1742
Db 405 aAlaGlyAlaThrGlyAlaAlaGlyThrGlyGlyCysAlaGlyGlyThrGlyAl 425
QY 1743 -----CGAGTGGCTGGCCAAAGCAGT---TGCGCGCGGCGGATCGTCTCTGGCGCT 1792
Db 425 aThrCysAlaThrAlaGlyGlyAlaGlyCysCysAlaGly--AlaGlyCysThrGlyAla 444
QY 1793 CCGCTCCCTCAGCGCGCTTACCGCGAGCTGATCCAGAGCGGGCTTCGACGTGGCT 1852
Db 445 GlyAlaThrGlyAla-----CysThrThrCysAla 454
QY 1853 GCATCCGCGCGGCGCACTCAGGCTTACATGGACCGGTCAACATGTACAGCTCTCTTACA 1912
Db 455 AlaAlaAlaCysAlaCysCysAlaGlyAlaGlyThrAlaCysThrThrGlyAlaAla 474
QY 1913 TGGCGCGCGGAGGCGG 1930
Db 475 GlyCysThrGlyGlyAla 480

RESULT 13

US-11-096-568A-19864
; Sequence 19864, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096.568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 19864

LENGTH: 766
TYPE: PRT
ORGANISM: Zea mays subsp. mays
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)-(766)
OTHER INFORMATION: Ceres Seq. ID no. 12375135
US-11-096-568A-19864

Alignment Scores:
Pred. No.: 2,5e-08 Length: 766
Score: 258.50 Matches: 192
Percent Similarity: 32.5% Conservative: 38
Best Local Similarity: 27.2% Mismatches: 234
Query Match: 7.1% Indels: 243
DB: 7 Gaps: 42

US-10-620-914-44 (1-1947) x US-11-096-568A-19864 (1-766)

QY 57 GGAGAGCTCAAGCTCAGCAGCATGAAGATGACTGACCGTTCTGCCCATATATGGTT 116
Db 191 GlyAspAlaGlyAlaHisGlnHisPro-----ProGlyLeuLeuArgHisProLeu 208
QY 117 CGCAGCAGAGGCGCATGATCA-----CGCTCTCGCTGGAGAGTTCTA 164
Db 209 ArgAspProArgGlyHisHisGluAlaAlaGlnHisArgCysGlnPro-----LeuPro 226
QY 165 CGGCGCCCA-----GGCGCTGCTTCTGCTGCCCGCTGGCGGAGCG 206
Db 227 AlaAlaProGlyHisHisArgValGlyArg-----ProGlyProAla 241
QY 207 CTC-----GAACCTCATCTCGGTGACCTGGTGGTGGCACTGGGGAGAATGTGCA 257
Db 242 LeuLeuHisArgArgProHisHisGlyProPro----- 252
QY 258 TATGATGGCTGATTACATCGACCTGGCGAAGTTCAAGTCCATCTACGTGGTCGACCTGTG 317
Db 253 -----GlnArgAlaGlyArg-----HisArgArgArgLys----- 262
QY 318 CCACCTCGTGTGCGAGGTGGCGAAGAGAGGCGAAGCGGCTGGAAGAATGTCCA 377
Db 263 -----GluGlyArgArgArgGlyValGlnAspArg 274
QY 378 GGTGTGGAGCGCGACGCTTGCCTCAATTTGGCGCCCTCGAGGG-----CACGCGAC 428
Db 275 HisArgGly-----ArgLeuLeuAlaGlnProGlnGlyGlyProArgHisArgGln 292
QY 429 GCTCATCACCTTCTCTACTGCTCAGATGATTCACCGTTCCACAACTCATCGACCA 488
Db 293 ArgHisValArgGlyLeuArgAlaArgGlyHisArgAspVal---ArgArgGlnArgPro 311
QY 489 GGCTGTGCTGTACTGTCCCAAGACGCGCTGGTGGCGGTGGCCGACTTCTACGT----- 542
Db 312 GlyArgProValGlyGlyProValArg-----ArgLeuLeuArgGlyHis 326
QY 543 GAGCGCAAGTACGACCTGCCCTCGCCAGATGCCCTGCTGGTGGCGCGCTTCTTCTGCGG 602
Db 327 GluArgGlnAlaArgValHisGlyProProAsp----- 337
QY 603 ATCGATCTTCGACATCGACAACTTGCATCGGCGCGCGCGCTTACCTTGGAGCA 662
Db 338 -----ProGlnAlaGluAlaProProGlyVal 346
QY 663 GAAGCTGGAGCGGTGGGAGCAGAACACCCAGGTTTCGATCCCTACGTGGCGGTGCT 722
Db 347 HisArgGlyArgGlyHisHisGlyAlaHisProGlyTrpGlnLeuLeuHisGluAla----- 365
QY 723 GCGCGCCCCCTACTACGTGTGGATTTGCCCGCTGCCAGCGTTGGCCACGCCCCGACGA 782
Db 366 -----GlyGlnGluGlyGluArgAlaGlyPro 374
QY 783 GGAGCGGTGGAGCGCGCCCATGTTCCGCGCCACCTTCTGTACAGCGAGTGTGGGA 842

QY 1636 -----CGGTCAAGTTGTCACACCGCCACTCTTGAAGGTGG----- 1601
Db 376 SerProArgSerGlyCysProArgGlySerLeuGlyValAlaProSerArgGly 395
QY 1600 -----CGA 1598
Db 396 CysGlyArgSerLeuArgGlyLeuProProGluAlaArgGlyAlaArgArgArg 415
QY 1597 AGGCGCGCTCGCGCAGGTAGGTGG----- 1574
Db 416 ArgProGlyGlyArgThrThrTyrProSerGluAsnArgGlnArgLysGlnLysLys 435
QY 1573 ---GGCACTTGTGGGCA---GGAACTTGGCGGTGAGGAGTTGTAGTAGTTCT 1520
Db 436 GluAlaSerCysArgSerArgGlySerProArgArgGlyGlnCys----- 450
QY 1519 GCTTGGCAGCTGCGAGTTCTCCGCGCAGCGCTCCATGTCGCGCGATGTAGTTCTCAA 1460
Db 451 -----ProThrGluProTrp-----GlyGln 457
QY 1459 TGGGATGCGTCCGCTT-----TGATCAGCGCGTACTGCTTGTCCCGGCACGC 1412
Db 458 GlyArgCysGlnProProArgSerProTrpGlnSerAlaLeuThrLysAlaGlnAlaSer 477
QY 1411 CGC-----CGCGAACCACAGCACCGGCTTGTGAAGACACCA 1373
Db 478 ArgMetProProGlyGlyGlyLeuArgGluAlaArgGlnAlaLysArgAlaGln 497
QY 1372 GGCTCAGCACTTGACGACAGCACACACCGGCTTGGGCT-----CGT 1328
Db 498 AsnSerArgGluArgAlaSerAlaLeuSerLeuTrpAlaTyrArgArgGlyGlyArg 517
QY 1327 TCTTCAAGTGTGATGATGATGTTGCTGCCACAGCGGCTGCTCTCTCCATTGGG 1268
Db 518 ArgGluArgAlaGly-----GlyAlaArgProAlaValTyr 529
QY 1267 GCGGCTTGGCAGCGCTGTGACGTTGCTTGGCCAGTCCAGCACCGCCAGGCACGTGCA 1208
Db 530 GlyArgThrHis-----ArgProArgArgProProValHis 541
QY 1207 GCACCCAGCAGCTTGCCTCCATGCGCCCTGCTAGTACAGCGCTGCTGGAGTACCAGA 1148
Db 542 SerAsnThrSerAlaGlyAspCys-----AlaGlyArgLysArg 554
QY 1147 GCGCTTGGACACAGAAAGTTGTGGCTTTCGACAGAGAGGGCGGAGCTTCTTCGT 1088
Db 555 GlyGlnHisSerArg-----ArgGlyProCysSerLeuArg 566
QY 1087 ACAGCTCTCAATGCGGGT-----GCACGCGCTCGCGCAACAGCT 1046
Db 567 LeuSerProCysSerLeuGlyLeuAspGlyIleTrpIleProHisProAspGlyArgGly 586
QY 1045 GCCACAGCTCTCAACTCCAGCTGCTGAATGGCA-----CCTTCTTCACTCCA--- 995
Db 587 GlyProGluTrpAlaSerSerAlaAspHisTyrProThrValProSerGlyAlaProThr 606
QY 994 -----GAAGCGCCGACT 983
Db 607 LeuProArgHisAlaGluHisAsnSerSerLeuProProLeuThrAlaGlnAlaSerThr 626
QY 982 GCGCGGGTTGACGTCCACCGACACCACTTGGCGCGCCCTCGCACCGAGGTTCAAGG 923
Db 627 AspThrGlyAlaValArgProAlaGlyProSerSerGlyThrSerThrAlaSerSerThr 646
QY 922 CAT---TGCACCGCGCTAGTACAGG-----TCAGCACCGCTGCTTGGGGTTGA 875
Db 647 HisSerCysLeuGluProThrSerGlyLeuArgArgSerProCysSerAlaAla 666
QY 874 TCTCATCACCTCCATAT----- 857
Db 667 AspSerHisSerProSerAsnThrLeuTrpIleSerGlyCysGlyAlaValArgGluSer 686
QY 856 -----CGGCTCGGGTCTCT-----CCACG 836

Db 687 GluLysSerGlyGlnAlaArgGlyAlaAlaValGlyValArgArgSerGlnGlyProLeu 706
QY 835 ACTCGGTGTACAGNAGS---TGGCGCGGACATGGCGCGCTCCACGCGCTCTCGT 779
Db 707 SerGlnSerLeuGlyArgAspThrProGlySerGlyAlaMetSerProValProArg 726
QY 778 ---GCAGGCGCTGCCAACCGCTGGCGCAGC---GCCCAATCCACACGCTAGTAGGGCGC 725
Db 727 CysAlaGlyLeuGlyThrAlaGlyHisGlyAlaGlySerSerArgAlaArgGlyArg 746
QY 724 GCA-----GCCACGCGCAGTAGG----- 707
Db 747 GlyGlySerGlyAsnAlaGlnAlaThrAlaGluGlyAlaGlyArgSerProAlaAspGly 766
QY 706 -----GGATCGAACCTGGGTGTTCTGCTCCACACGCGCTCCAGCTTCT 662
Db 767 ThrAlaGluGlyTrpGlyGlyHisProAsp---GluAlaAlaLeuHisSerGlnGln 785
QY 661 GCTCAGGTAGGCGC-----GGCGCTCGGGCGCGATGTCAATGT 623
Db 786 AlaGlyGlyProArgProSerTyrThrValSerGlySerGlyArgGlyHisSerProAla 805
QY 622 TGT-----CGATGTGGAAGATCGATCCCGCAGAAAGCGCGCACCGGCGCATCT 572
Db 806 CysAlaProLeuGlnProAlaArgAlaAlaAlaArgAlaAlaValProAlaAlaThr 825
QY 571 GGCGCAGG-----GCAGGTCTGACTTGC----- 548
Db 826 ProAlaGlyValAlaValAlaLeuGlyProArgGlyGlyAlaAlaGlyLysLeuCysTyr 845
QY 547 -----CGCTCAGTAAAGTCCG-----CAACGC 524
Db 846 AlaTrpGlyHisLeuValCysProAlaGlyAspArgAlaAlaIleProCysPheGlnArg 865
QY 523 CCACAGCGCTTGGGACAGGTACGAGCAGCTGCTGATCGATCGTGTGGACCGTG 464
Db 866 GlyProGlyAlaProGlyProGlyThrGlyGlyProGlySer----- 879
QY 463 GAATCATCGTGAGCGAGTAGGAGAAGGTGATGAGCGTCCGGTGC-----CCTCAGGG 410
Db 880 -----GlyGlyGluCysHisLeuProGluGly 888
QY 409 GCGCAATGGCAAGCGCTCGGCTCCACGACCTCGACAT-----TCTTCCAGC 362
Db 889 Ala-----GlyTyrLeuThrAlaAlaGlyAlaGly 898
QY 361 CTTGGCTTCCGCTTCTTCTTGGCCACCTCGCACAGCG-----AGTGGACA--- 314
Db 899 GluTrp-----ProProSerGlyProAlaAlaThrAlaLeuLeuProAlaGlyThrMet 916
QY 313 -----GGTCGACACCGTAGATGGACTTGAACCTCCGCA---GGTCGATGT 272
Db 917 GlyGlyLeuCysSerAlaGlyArgSerTrpAlaGlySerGlyGlyCysAlaGlyCys 936
QY 271 AATCAGCCCATCATCATGACATCTCCCGAGTCCGACAC----- 233
Db 937 ThrGlyProAlaAlaGlyProArgAlaGlnCysArgHisLeuProGlyAspAlaGlyPro 956
QY 232 -----CCAGTCAACCCAGATGAGTGTTCGAGCGCTCGGCCAGCGCGGCGACAAAG 182
Db 957 GlyProGlyProGlyGlnPro-----SerSerProAlaArgMetGlyProLeuProGly 974
QY 181 CAGCGGCTGGGCGCGTAGAGCTCTCCAGGCGAGCAG---CGTGTATCATCGCCCTTCT 125
Db 975 ProLeuProArgAlaArgGluGluAspProAlaThrProGlyArgGlyGlyGluProThr 994
QY 124 TGCTGCCGAACCATATGTC----- 104
Db 995 GlyLeuProThrThrAlaGlyArgArgCysGlnClnTrpArgGlyProValGlyAlaPro 1014
QY 103 -----GCAGAACGGTCAAGTCTATCTTCTATGCTGC 74
||| :|||

Search completed: March 14, 2006, 01:53:18
Job time : 49 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: March 14, 2006, 01:57:16 ; Search time 26 Seconds
(without alignments)
2398.018 Million cell updates/sec

Title: US-10-620-914-45

Perfect score: 3463

Sequence: 1 MGSGRDRPASVTKNFSLE.....RVNYSFFYARRKGAOKDN 648

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 80.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	326.5	9.4	416	2 A97614	hypothetical prote
2	326.5	9.4	416	2 AG2836	conserved hypothet
3	120	3.5	449	2 S07714	T64 protein precu
4	120	3.5	451	2 I50131	clusterin - quail
5	112.5	3.2	2285	1 G02434	DNA-directed DNA p
6	106.5	3.1	180	2 B69524	rRNA (adenine-N6) -
7	106	3.1	472	2 I59087	ISG-K54 - human
8	106	3.1	2039	2 S64540	probable calcium c
9	105	3.0	234	2 T48892	2-heptaprenyl-1,4-
10	104.5	3.0	664	2 F83376	conserved hypothet
11	102.5	3.0	465	2 AF1858	hypothetical prote
12	102.5	3.0	493	2 T09919	hypothetical prote
13	102	2.9	656	2 A48439	dnak-type molecula
14	101.5	2.9	663	2 E90689	cytochrome o ubiqu
15	101.5	2.9	663	2 B42226	bo-type ubiquinol
16	101.5	2.9	663	2 A85540	bo-type ubiquinol
17	99.5	2.9	663	2 AE0557	cytochrome o ubiq
18	98.5	2.8	552	2 E72283	alpha-galactosidas
19	98.5	2.8	673	2 T31610	hypothetical prote
20	98.5	2.8	1336	2 T02736	probable SCARECROW
21	98.5	2.8	2890	2 B71846	dna-directed RNA p
22	98	2.8	993	2 A10669	probable virulence
23	97.5	2.8	280	2 T36376	hypothetical prote
24	97.5	2.8	662	2 B75544	arginine decarboxy
25	97	2.8	239	2 AH2836	O-methyltransferas
26	97	2.8	239	2 B97614	hypothetical prote
27	97	2.8	363	2 G82376	recP protein VC001
28	97	2.8	366	2 T33885	hypothetical prote
29	96.5	2.8	152	2 AE0356	transposase, IS154

30	96.5	2.8	152	2 AC0003	transposase, IS154
31	96.5	2.8	152	2 AE0195	transposase, IS154
32	96.5	2.8	152	2 AC0127	transposase, IS154
33	96.5	2.8	152	2 AE0349	transposase, IS154
34	96.5	2.8	152	2 AF0154	transposase, IS154
35	96.5	2.8	189	2 H70323	hypothetical prote
36	96.5	2.8	471	2 E83130	conserved hypothet
37	96	2.8	434	2 D72353	lipopolysaccharide
38	96	2.8	472	2 S29243	interstitial colla
39	95.5	2.8	426	2 B53652	rhannosyltransfera
40	95.5	2.8	448	2 S45112	transcription fact
41	95.5	2.8	2890	2 P64669	DNA-directed RNA p
42	95	2.7	252	1 B48653	hypothetical prote
43	94.5	2.7	152	2 AF0012	transposase, IS154
44	94.5	2.7	169	2 T14928	transposase - Yers
45	94.5	2.7	2470	2 S57085	1-phosphatidylinos

ALIGNMENTS

RESULT 1

A97614
hypothetical protein AGR_C_3843 [imported] - Agrobacterium tumefaciens (strain C58, Cerc
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 09-Jul-2004
C:Accession: A97614
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: A97614
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-416 <KOR>
A:Cross-references: UNIPROT:Q8UDK6; UNIPARC:UPI00000D1DB1; GB:AE007869; PIDN:AAK97866.1;
C:Genetics:
A:Gene: AGR_C_3843
A:Map position: circular chromosome

Query Match 9.4%; Score 326.5; DB 2; Length 416;
Best Local Similarity 27.8%; Pred. No. 2.6e-18;
Matches 111; Conservative 69; Mismatches 158; Indels 61; Gaps 16;

QY	275	LYTQSWEDPEPMVEINPKDVLTLTSGGNALNLLVQAGQVVSVCNPAQSALEL	334
DB	44	VTPQIWEDEPDMEAMELGEGHRIVTIGSGCNMLAYLSRNPASIDVDLNPHEIALNKL	103
QY	335	KKVAIQQL-EFEDVMQLFGEGVHPRIEELYEKKLAPFLSQTSHNFSK-----RLWYF	386
DB	104	KLAAPRHLPAHQDVVRHFGAGTRSNVGVYDFRFAEHLDAATTKAYWSKRTLSGRRRISVF	163
QY	387	QHGLYQQGNGKLCWVLQCLAVVLGKTVKRLANAPTEBQRRLWDSNMLHFVNGPK	446
DB	164	DRNIYRTGLLGRFIGAGHIMARLHGKLT--EMAKTRTLDQRPQFDSKVAPLF-----DK	217
QY	447	PLV-MLFKVFSVLNFKAVLMFGGVPCKQY---ALIKADGIPINAYLTARTWDGVAENS	502
DB	218	PVVRMLTKRKSLL-----FGLGIPPRQVDELASLSDG--TVASVLKERLEKACNP	267
QY	503	HYRKQNYFYINCLTKGFLRDN---CPTYLREAAFATLKSQGVVDNLTVSTNPFMEELK---	556
DB	268	PL-SDNYFAWQAFARRYPEPEHGALPAYLKPEYKIRNNTA-RVAVHATYTELSSRP	325
QY	557	ARTYTKVILMDHVDLMDPVAELAECLAKVAPGGIVWRGAS-----LSPPYAEL	608
DB	326	ANGVDRIYLLDAQDWMVDVQLNELWSQISRTAASGARVIFRTAAEKSVIEGRLLSP-----	380
QY	609	IQKAGFDVR-----CIRRATQ-GYMDRVNMYSSFFYMARR	641
DB	381	-----DIRNQWYLEERSNELNAMDRAIYGGFHIYQR	413

RESULT 2
AG2836
conserved hypothetical protein Atu2119 [imported] - Agrobacterium tumefaciens (strain C58)
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
C;Accession: AG2836
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monke, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.; Karp, P.; Romero, P.; Zhang, S.
Science 284, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, S.E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; MUID:21608550; PMID:11743193
A;Accession: AG2836
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-416 <KUR>
A;Cross-references: UNIPROT:Q8UDK6; UNIPARC:UPI00000D1DB1; GB:AE008688; PIDN:AAL43109.1;
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: Atu2119
A;Map position: circular chromosome

Query Match 9.4%; Score 326.5; DB 2; Length 416;
Best Local Similarity 27.8%; Pred. No. 2.6e-18;
Matches 111; Conservative 69; Mismatches 158; Indels 61; Gaps 16;

QY 275 LYTQSWEDPEPDMEVMEINPKDTVLTLTSGGCNALLVQAGQVSVDCNPAQSALLEL 334
DB 44 VYPQWEDPEIDMEAMELGHGHRIVITIGSGCNLAYLSRNPASIDVVDLNPHHIALNKL 103

QY 335 KVAIQOL-EPEDVQVLPFGVHPRIEELYEKKLAPLSQTSNFWSK-----RLWVF 386
DB 104 KLAARHLPAHQDVVRHFGFRAGTRSNVGVDFRFAEHLDAATRAYWSKRTLSRRRISVF 163

QY 387 QHGLYVQGGWGLCWVLCVLAIVGLGKTVKRLANAPTMEEQRLWDSNMLIHFKVNGPK 446
DB 164 DRNIYRTGLGRIGAGHINARLHGKLT--EMAKRTLDQEQFDSKVAPLF-----DK 217

QY 447 PLV-WLFVKFVSVLNFKAFLWFGGVPKGQY---ALIKADGIPNIENYIARTMDGVAENS 502
DB 218 PVVRWLTTRKSSL-----FGLGIPRQYDELASLSDG-TVASVLKERLEKLACNF 267

QY 503 HVKQNYFYNYCLTGKFLRN---CPTYLREAPATILKSGVDNLTVSTNFFWEELK--- 556
DB 268 PL-SDNYFAWQAFRRYPPEHGPALPAYLKPPEYKIRNNTA-RVAVHHAITYTELLSRKP 325

QY 557 ARTYTKVILMDHVDLMDMPVANLAELAKQVAPGGIVIVWSAS-----LSPPYAEL 608
DB 326 ANGVDRIYLLDAQDMTVDQVNLWNSQISRTASGARVIFRTAAKSVIEGRISP----- 380

QY 609 IQKAGFDVR-----CIRRTAQ-GYMDRVNMYSSFFYMR 641
DB 381 -----DIRNQWVYLEERSNELNADRSALYGGFHIYQR 413

RESULT 3
S07714
T64 protein precursor - Japanese quail
C;Species: Coturnix coturnix japonica (Japanese quail)
C;Date: 29-Jan-1993 #sequence_revision 29-Jan-1993 #text_change 09-Jul-2004
C;Accession: S07714
R;Michel, D.; Gillet, G.; Volovitch, M.; Pessac, B.; Calothy, G.; Brun, G.
Oncogene Res. 4, 127-136, 1989
A;Title: Expression of a novel gene encoding a 51.5 kD precursor protein is induced by d
A;Reference number: S07714; MUID:89239492; PMID:2541393
A;Accession: S07714
A;Molecule type: mRNA
A;Residues: 1-449 <MTC>
A;Cross-references: UNIPROT:P14018; UNIPARC:UPI00001711C7; EMBL:X15825; NID:G62594; PIDN
C;Superfamily: clusterin

F;1-18/Domain: signal sequence #status predicted <SIG>
F;19-449/Product: T64 protein #status predicted <MAT>

Query Match 3.5%; Score 120; DB 2; Length 449;
Best Local Similarity 22.4%; Pred. No. 0.13;
Matches 94; Conservative 55; Mismatches 157; Indels 114; Gaps 23;

QY 105 LCHSLCEVAKKAKAGKWKNVQVEADACQAFAPPEGTATLTITSYSLTMTIPPHNVIDQA 164
DB 66 MLHTLEETKKKEEA-----VKLALEKEQLAEKQEVNCNETMLSLWEECKPCLKHTCMRV 120

QY 165 CSYL--SQDLVGV--VADF-----YVSGKY--DLPLRQMPWRRFFRWSIFDIDNI 209
DB 121 YSKMCHSGSLVGRQLEEFLENRSSPSIWNNGRIDDLDRQQRERF-----EDL 172

QY 210 DIGPERRAYLEQKLERVWEQNTQGSIPYVPLRAPYVWIGR--LPSVG--HALHEERVE 265
DB 173 E---ERFGLMEDGVEDIFQDSTQLYGPAFFPTPTFGGFRFAFPVQVRVHLVPRRLS 229

QY 266 RP-----PM-----FPPTFLYTQ-----SWEDPEPDMEVMEIN-PKDTV---L 299

DB 230 RELHPFFQHPMGFHLFQLEFEMTOHMLDGGHAWEHPLGGPATESRNFSTDRMVCRI 289

QY 300 TLTSGGCNALLVQAGQVSVDC---NPAQSALLEKKAIVQLEFEDVQVLPFGVH 356
DB 290 RNSAGCLRMRDECEKREILAVDCSTQDPVQSOLRE-----QFEDALRL-----A 335

QY 357 PRIELIYEKKLAPF-----LSQTSNHF-WSKRLVYFQHGKLYQGGWGLCWVLOC 405
DB 336 ERFTTRYDLLSAFOAEMLNTSLLDQLNRQFGWVSRLL-----GNLTQGNDG---FLQV 386

QY 406 LAVVLGLGKTVKRLANAPTMEEQRLWDSNMLIHFKVNGPKPLV-----WLFVKFVSVL 459
DB 387 TTV-----FSKTPNLEDPSPADTQVTVQLFSEPLSLTVPGDISWDDPRFMEIV 436

RESULT 4
I50131
clusterin - quail
C;Species: Coturnix coturnix (quail)
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 13-Aug-1999
C;Accession: I50131
R;Michel, D.; Chatelain, G.; Herault, Y.; Brun, G.
Eur. J. Biochem. 229, 215-223, 1995
A;Title: The expression of the avian clusterin gene can be driven by two alternative pro
A;Reference number: I50131; MUID:95262670; PMID:7744033
A;Accession: I50131
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-451 <MIC>
A;Cross-references: UNIPARC:UPI0000127BA5; EMBL:X80760; NID:G520629; PIDN:CAA56733.1; PFI
C;Genetics:
A;Gene: T64
A;Introns: 29/1; 78/3; 135/3; 275/1; 313/1; 388/3; 450/2
C;Superfamily: clusterin

Query Match 3.5%; Score 120; DB 2; Length 451;
Best Local Similarity 22.4%; Pred. No. 0.13;
Matches 94; Conservative 55; Mismatches 157; Indels 114; Gaps 23;

QY 105 LCHSLCEVAKKAKAGKWKNVQVEADACQAFAPPEGTATLTITSYSLTMTIPPHNVIDQA 164
DB 66 MLHTLEETKKKEEA-----VKLALEKEQLAEKQEVNCNETMLSLWEECKPCLKHTCMRV 120

QY 165 CSYL--SQDLVGV--VADF-----YVSGKY--DLPLRQMPWRRFFRWSIFDIDNI 209
DB 121 YSKMCHSGSLVGRQLEEFLENRSSPSIWNNGRIDDLDRQQRERF-----EDL 172

QY 210 DIGPERRAYLEQKLERVWEQNTQGSIPYVPLRAPYVWIGR--LPSVG--HALHEERVE 265
DB 173 E---ERFGLMEDGVEDIFQDSTQLYGPAFFPTPTFGGFRFAFPVQVRVHLVPRRLS 229

QY 266 RP-----PM-----FPPTFLYTQ-----SWEDPEPDMEVMEIN-PKDTV---L 299

Db 230 RELHPFFQHPHGHFRLFQPLFEMTQHMLDGGHAWHEHPLGSPATSRNFSFTRMVCREI 289
 QY 300 TLTSGGNCALNLVQAGGVSDC--NPAQSALLELKKVAIQOLEPEFEDVWOLFCEGVH 356
 Db 290 RNSACGLMRDCEKREILAVDCSQTDVQSQLRE-----QFEDALRL-----A 335
 QY 357 PRIEELYEKKLAPF-----LSQTSNPF-WSKRLMYFOHGLYYQGMGKLCWVLOC 405
 Db 336 ERFTREYDOLLSAFOAEMLNTSLLDQLNRFQGW SRL-----GNLTQNDG----FLQV 386
 QY 406 LAVVLGLGKTVKLANAPTMEEQRRLWDSNMLTHFVNGPKPLV-----WLFVKFVSLV 459
 Db 387 TTV-----FSKTPNLEDSAPADQTVTVOLFDSBPLSLTVPGDISWDDPRFMEIV 436

RESULT 5
 G02434
 DNA-directed DNA polymerase (EC 2.7.7.7) epsilon catalytic chain A [validated] - human
 C:Species: Homo sapiens (man)
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 20-Apr-2000
 C:Accession: G02434; A46692
 R:Asahara, H.; Goldsmith, J.S.; Lee, E.; Linn, S.
 submitted to the EMBL Data Library, February 1996
 A:Reference number: H01252
 A:Accession: G02434
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: mRNA
 A:Residues: 1-2285 <ASA>
 A:Cross-references: UNIPARC:UPI000016A2D7; EMBL:U49356; NID:g1206034; PIDN:AAA90924.1; F
 J. Biol. Chem. 268, 10238-10245, 1993
 A:Title: Molecular cloning of the cDNA for the catalytic subunit of human DNA polymerase
 A:Reference number: A46692; MUID:93252906; PMID:8486689
 A:Accession: A46692
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-68,96-441,'T',444,'AGS',448-1081,'NA',1084-1297,'E',1299-1371,'P',1373-151
 A:Cross-references: UNIPARC:UPI000016A096; EMBL:S60080; GB:L09561; NID:G303156; PIDN:AA
 A:Experimental source: HeLa cells
 A:Note: sequence inconsistent with the nucleotide translation
 A:Note: sequence extracted from NCBI backbone (NCBIN:131270, NCBI:P:131271)
 C:Genetics:
 A:Gene: GDB:POLE
 A:Cross-references: GDB:129548; OMIM:174762
 A:Map position: 12q24.3-12q24.3
 C:Superfamily: DNA-directed DNA polymerase II
 C:Keywords: DNA binding; nucleotidyltransferase

Query Match 3.2%; Score 112.5; DB 1; Length 2285;
 Best Local Similarity 19.4%; Pred. No. 4.9;
 Matches 135; Conservative 83; Mismatches 199; Indels 279; Gaps 34;
 QY 107 HSLCEVAKKAKAGHKNVQVEADACQAPPEGTATLTFYS----- 150
 Db 422 HNL-----KAAASKGLGVDVPELDPEDMCRMA-TEQPQTATYSVSDAVATYLYIMKYVHPF 476
 QY 151 -----LTMIPPPH-NVI-----DQACSVLSQDGLGVAD 178
 Db 477 IFALCTIIPMEPDEVLRKSGTGILCEALLMVQAFHANIIFPNKQEQBFNKLTDGHHVLDSE 536
 QY 179 FTVSG-----KYDLPRLQMPWSRRFFWRSIFPDIDNIDIGPERRAYLEQKLERVWE 228
 Db 537 TVVGHVHALESQVFRSDIPCR-----FRMNPAAFDLLQVREK----- 575
 QY 229 QNTQSGIPVFWLRLAPYYWIGRLPSVGHALHHE----- 262
 Db 576 -----TLRHALLEEKVPVEQVTNFEVCEDEIKSKLASLK 610
 QY 263 -----RVREP-----PMFPPTFLYTQSWEDPEPDMWEINPKDVTLTLSGCN--- 307
 Db 611 DVPSRIECPLIYHLOVGAMYPNLTNR-----LQPSAMVDEATCAACDFNK 657

QY 308 -----ALNLLVQAGGVSDCNPAQSALLELKKVAIQOLEPEFEDVWOLFCEGVHPRIEEL 362
 Db 658 PGANCQKMAWQWRGSEFM-----PASRS--EVHRIQ-HOLESEKFPPLPEGPAPARAFHEL 709
 QY 363 -----YEK-KLAPF-----LSQTSNPFWSKRLMYFOHGLYYQGG 395
 Db 710 SREBQAKYEKRRLADYCRKAYKKIHITKVEERLTTTCORENSFYVDTVAFRRDRRYEFKG 769
 QY 396 MGKLCWVLOC LAVVLGLGKTVKLANAPTMEEQRRLWDSNMLTH-----FVKNGP 445
 Db 770 LHKVWKKKLSAAVEGDAAEVKCKN--MEV--LYDSIQLAHKCILNSFYGVYVWRKA 823
 QY 446 KPLVWL FVKFVSLV LFNKAVLWFGGVGPKQKVALIKADGIPENYIARTWDG---VAENS 502
 Db 824 R---MYSEMAGIVCFT-----GANIITQARELIEQIGRPLE-----LDTDGIWCVLPS 870
 QY 503 HVRKQNYFYVNCITGKFLRDCNCTYLREAAFPATLKSGVVDNLTVSTNF-----PMBELKAR 558
 Db 871 F--PENFVFKT-----TNVKPKVTISYPGAMLNIMVKEGFTNDQYQELAEPS 916
 QY 559 TYTKVILMDHVDWLD-----MPVANLEAECLAKQVAGGIVIVRSASLSPPYAELI 609
 Db 917 SLTYVTRSENSIFFEVDGPLYLAMILPASKEEGKCLKRYA-----VFNEDGSL-----AEL- 967
 QY 610 QKAGFDVRCIRATQGYMDRVNMY-SSFYMARCKGA 644
 Db 968 --KGPEVK-----RRGELQLIKIFQSSVFEAFKGS 996

RESULT 6
 B69524
 rRNA (adenine-N6)-methyltransferase homolog - Archaeoglobus fulgidus
 C:Species: Archaeoglobus fulgidus
 C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 12-Jul-2004
 C:Accession: B69524
 R:Klenk, H.P.; Clayton, R.A.; Tomb, J.P.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson,
 ; Fleisemann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.;
 Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
 Nature 390, 364-370, 1997
 A:Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.
 Smith, H.O.; Woese, C.R.; Venter, J.C.
 A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
 A:Reference number: A69250; MUID:98049343; PMID:9389475
 A:Accession: B69524
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-180 <KLE>
 A:Cross-references: UNIPROT:O28089; UNIPARC:UPI0000056A3B; GB:AE000954; GB:AE000782; NID
 C:Superfamily: spore germination protein C2

Query Match 3.1%; Score 106.5; DB 2; Length 180;
 Best Local Similarity 22.3%; Pred. No. 0.43;
 Matches 48; Conservative 38; Mismatches 64; Indels 65; Gaps 14;
 QY 34 LRHMWFGSKGD--DHAARLESFYGQAAFAARLAERSNLWIWDLGGGTGENVMDY 91
 Db 1 MKHK-FDPKHAHLDSEWRKIFPPEKVEFTESLRPKQVLPF-DVGAGTG-----Y 50
 QY 92 --IDLAK-FKSYVVDLCHSLCEVAKKAKAGKNVQVVEADACAPAPEGTATLITPS 148
 Db 51 LTIPLARVFKKYAYEISEMAEVLRLRVEEGLNIGIIVSE---KPPE-----VDPR 101
 QY 149 YSLTWIPPFHNVIDQACSVLSQDGLGVADFVSGKYDLPRLQMPWSRR-----PFWRS 202
 Db 102 VDVL---FSNVLHE-----MDNPEYLEWASADYVVAEWKK 137
 QY 203 IPDIDNIDIGP--ERRAYLEQ-----KLERVWEQN 230
 Db 138 ---EKTEFGPPVEERLSLEENLSKMKLVKSEN 168

RESULT 7
 I59087

QY 143 TLITFSYSLTWIPPHNV-----IDQACSYL--SQDG----- 172
 Db 238 QIAAMGFDVLIYFPPIHPIGTRHKGRRNSLRAEAGDPCSPYAIGSEEGHGAHPELGDR 297
 QY 173 -----LVGV-----ADFYVSGKYDLP-LRQMP-WSRPFPPWR-----SI 203
 Db 298 EDFRLLIVAREHGMELALDFAQCSDPHWRLEHFGW--FANRPDGLRYAENPPPKY 354
 QY 204 FDIDNIDIGERRAYLEQKLERVWEQNTQGSIPYV-----PWLRL-APYVWI-- 249
 Db 355 EDIVNVDF-----YABQALPSLWEALRDVVILGWVEQVTLFRVDNPHTKPLPFWEMLIA 408
 QY 250 ---GRPLSVCHALHEERVERPPM-----PPTFFLY-----TQSW 280
 Db 409 EVRGRHPQV--IFLSEAFTRPAMMARLGKVGFSQSYTYFTRWRNDKQELAEYFAELNQPPW 466
 QY 281 ED---PEPDMVMEINPKDVTLTISGGCNALNLLVOGAG-----QVVSVDNCNPAQS 329
 Db 467 RDCYRNPFFVNTPDINPWFLOSRGRPGLIRAAATWGSGLWGMYSGFELCEAAALPGKE 526
 QY 330 ALLELKVAIQOLEFEDVWOLFGEVHPRIEELYEK--LAPPLSQTSNHFWSKRLWYF 386
 Db 527 EYLDSEKYQLRPDYQAPGNIVAB--IARLNIRRENPAQTHLGLGFOAYNAWNDRILYF 583

RESULT 11
 AF1858
 A:Title: hypothetical protein all0415 [imported] - Nostoc sp. (strain PCC 7120)
 C:Species: Nostoc sp. PCC 7120
 A:Note: Nostoc sp. Strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7122
 C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
 C:Accession: AF1858
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.
 C;Mol. Cell. Physiol. 187: 1-10 (1997)
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacter
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yasuda, M.;
 DNA Res. 8, 205-213, 2001
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacter
 A:Reference number: AB1807; MUID:21595285; PMID:11759840
 A:Accession: AF1858
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-465 <KUR>
 A:Cross-references: UNIPROT:Q8VZP2, UNIPARC:UPI00000CDD3B, GB:BA000019; PIDN:BA
 A:Experimental source: strain PCC 7120
 C:Genetics:
 A:Gene: all0415
 C:Superfamily: Streptomyces coelicolor hypothetical protein SCE29.06c

Query Match 3.0%; Score 102.5; DB 2; Length 465;
 Best Local Similarity 19.7%; Pred. No. 3.4;
 Matches 75; Conservative 50; Mismatches 150; Indels 105; Gaps 18

QY 38 WFGSKGDDHAARLESFYGPQAAFAARLAERSNL---IWVDLG-----GGTGENVDM 87
 Db 109 WYNYSGVDSVTPRPRYPADGRKLPIYQENLRSLRLQIDLQGFPLFNWGPNTSIRS 168
 QY 88 MADYIDLAKFKS-----IYVDLCHSLCEVAKKAK-AKGMKNQVQVEADACQFAP 137
 Db 169 TQMIASAKFLGDCDPTLTVLPHLDYCLQKYGTDITKAKLQELDAICSDLIQFVE 228
 QY 138 PEGTATLITFSYSLTWI-PPFH-----NVIDQACSYLSQDGLVGVA 177
 Db 229 NRGAQVILVSEYGITVSQPIHINRVLRHGLLTIRBELGRELIDAGAKAFADHQA 288
 QY 178 DFFVSGKYDILPLQMPHSRRFPFWSIFIDINID-----IGPERRAYLEQKLERVWEQNTQ 232
 Db 289 HVTYNDPPYIP--QV-----RSL--LENIDGIADVLDITQKSYHYHNLHNSRAGELIAV 336
 QY 233 GSIPYYPWLRAPVYVWIG--RLPSVGHALHEERVERPMPPTFLYLTQSWEDPE---PDM 287
 Db 337 SQ-----PDWFTYYLWLLDRAPDFAKTVD---IHRKEGYDPVELFL-----DPEIKLPQV 385
 QY 288 EV-----MEINPKDVTLTISGGCNALNLLVOGAGQVVSVDNCNPAQSALLE 333

Db 386 KIATKLQKKGPRYLMDIPLDAELVKSHGC-----LPPSSQSPFLLI 430
QY 334 LKKVAI---QQLEFEDVWQL 350
Db 431 TQSHLFDSTAIGSATDYQL 450

RESULT 12
T09919
Hypothetical protein T16L4.70 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 09-Jul-2004
A;Accession: T09919
R;Bevan, M.; Rose, M.; Hempel, S.; Entian, K.D.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.
submitted to the Protein Sequence Database, June 1999
A;Reference number: Z16897
A;Accession: T09919
A;Molecule type: DNA
A;Residues: 1-493 <BEV>
A;Cross-references: UNIPROT:Q9SU89; UNIPARC:UPI000000AC268; EMBL:AL079344; GSPDB:GN00062;
A;Experimental source: cultivar Columbia; BAC clone T16L4
C;Genetics:
A;Gene: ATSP:T16L4.70
A;Map position: 4
A;Introns: 65/2
C;Superfamily: Arabidopsis thaliana hypothetical protein T16L4.70

Query Match 3.0%; Score 102.5; DB 2; Length 493;
Best Local Similarity 20.4%; Pred. No. 3.7;
Matches 102; Conservative 65; Mismatches 171; Indels 161; Gaps 24;

QY 176 VADPYSGKDYDLPLRQMPSPRRPFWRSIFIDNIDIGPERRAYLEQKLERVWEQNTQGS 235
Db 111 VDFWVQORDAHLNLDRIIP-KPKFEWISNLDLSAD-----EDSMKQVFD--SV 155
QY 236 PYVPWLRAPYVYVIGRLPSVGHALHEERVERPMPPT-----FLYTQSW----- 280
Db 156 P--DWLSE-----KLGSAGTIL-----PWLPSVCDVDSEMLVVDVSWNGKEITQI 198
QY 281 -EDPEPDME--VMEINPKDTVLTSGGCNALNLVQ-----GAGQVSVVDCNPAQSALL 332
Db 199 KEDMEEDQREVVRNIDHTMTVGLQAHDEMAVNLRAQISNFESTSEVLTIL-CNDIRRLCL 257
QY 333 ELKVAIQOLEFEDVWQLFQEGVHPRIEELYEKLAFLSQTSHNFSKELWTFQHGLYY 392
Db 258 EKGDSLRVLALTEPMNA-----DDETAALVSSLHN----- 289
QY 393 QGGMGKLCWVLCQCLAVVLGLGKTVKRLANAPTMBEQRRLWDSNNMLHFVKNGPKPLVWLF 452
Db 290 -----GSEEBELGHPSHVLCIV-----L 308
QY 453 VKFVSL-----VLFNKAVLWFGGVGPKQYALI-----KADGPIPIENYIARTWDGV-AE 500
Db 309 PKFLSLEKSASRVLSSTIEFKIHQRAAEYALVPLLRKEG--INNFCIEVISRVLKE 366
QY 501 NSHVKQNYFYNNCLTGKFLRDNCPYLREAAFYTL---KSGVVDNLTVSTNFWMEELKA 557
Db 367 CLHLGQISGFCQKLLCGR-----TQERRFMFLPCHRDLTISDELTWNEYLFILFQNI 417
QY 558 RYTKVILMDHVDMDMPVANELACLAKOVAPGGIVWRSASISPPYA-----ELIQKAG 613
Db 418 LIHGIPISQSDVCLVSKV-QELAERYSKSLKFGNFIHLHTAKCAPMLOAHKQLIIE--- 473
QY 614 FVRCIR-ATQGYMDRN 631
Db 474 -SVKCTNSLVTKSILSKIN 491

RESULT 13

A48439
dnak-type molecular chaperone Hsp70 - Entamoeba histolytica
N;Alternate names: heat shock protein Hsp70

C;Species: Entamoeba histolytica
C;Date: 01-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 31-Dec-2004
C;Accession: A48439
R;Ortner, S.; Plaimauer, B.; Binder, M.; Wiedermann, G.; Scheiner, O.; Duchene, M.
Mol. Biochem. Parasitol. 54, 175-183, 1992
A;Title: Humoral immune response against a 70-kilodalton heat shock protein of Entamoeba
A;Reference number: A48439; MUID:93063033; PMID:1435858
A;Accession: A48439
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-656 <ORT>
A;Cross-references: UNIPROT:Q24842; UNIPARC:UPI0000078A72; GB:M84652; NID:g158955; PIDN:
A;Experimental source: SFL-3, trophozoites
A;Note: sequence extracted from NCBI backbone (NCBIN:117218, NCBIP:117219)
C;Function:
A;Description: involved in protein folding and assembling/disassembling of protein complex
C;Superfamily: bcr protein
C;Keywords: ATP; molecular chaperone

Query Match 2.9%; Score 102; DB 2; Length 656;
Best Local Similarity 19.0%; Pred. No. 6;
Matches 121; Conservative 64; Mismatches 175; Indels 276; Gaps 27;

QY 57 POAAFAARLAERS-----NLIWVDLGGGT-----GENVDM 87
Db 179 PTAALAYGLDKKSDREKNVLPDLGGGTDFVSLLAIDDGVEFKASNGDTHLGGHDFDN 238
QY 88 MADYIDLAKFKSIYVVDL-----CHSLCEVAKKAKGKWKVQVVEADACQAPP 138
Db 239 RLNVNFIABPKRKKKDISGNARAVRLRTACRAKRTLSAATANIEV----- 287
QY 139 EGTATLITFSYSLTMTIPPFHNVIDQACSYLSQDGLVGADFYVSGKYDLPRLQMPWSRRF 198
Db 288 -----DQLFDGIDFYTS----- 299
QY 199 FWSIFDINID-----IGERRAYLEQKLERVWEQNTQGSIPYVWLRAPYVWIG--- 250
Db 300 ITRARPEELNIDLFKSTIGPVERVLQDAKLK-----GSIDDV-----VLGGST 344
QY 251 RLPSVGHALHEERVERPMPPTFLYTQSWEDPEPMEVMEINPKDTV-----LT 300
Db 345 RIPKVVQLLOD-----PFNGKEPN---KSNPDEAVAYGAQAVALT 384
QY 301 LTSGGCNALNLVQAGQVSVVD--CNPAQSALLLEKVAIQOLEFEDVWQLFQEGVHPRI 359
Db 385 GTGKATEDVLLLDVAPLTIGIETAGGVMTALI-----PRN 420
QY 360 EELYEKLAFLSQTSHNFSKELWTFQHGLYYQGGMGKLCWVLCQCLAVVLGLGKTVKRL 419
Db 421 STIPAKKSQVFSYADN-----QPGVLIQVFEGEASMTNHC----- 456
QY 420 ANAPTMBEQRRLWDSNNMLHFVKNG--PKPLVWL FVKFVSLVFNKAVLWFGGVGPKQY 477
Db 457 -----NLLGKFELTGTPAP-----RGVPOIEV 479
QY 478 AL-IKADGPIPIENYIARTWDGVAENSHVRKQNYFYNNCLTGKFLRDNCPYLREA-AFAT 535
Db 480 TPDIDANGI-----LNVSAEDKTGKKNTITNDKGRLSKEQIDRWVAEAEKFA 530
QY 536 ----LKSGV-----VNLTIVS-----TNFMELKARTYTKV--ILMDHVDWLD--MPVA 577
Db 531 EDDKMKQVRBAKNLENFCYSVKNTLSEQPADKIAAEDKTTIENIVKETLWDINNNQNAS 590
QY 578 NELAECLAKOVAPGGIVWRSASISPPYAEILQKAG 613
Db 591 TBYDNNMKKEV-----EGKVQPIFTKLYQOAG 617

RESULT 14

E90689
Cytochrome o ubiquinol oxidase subunit I [imported] - Escherichia coli (strain O157:H7,
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004

C;Accession: E90589
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genome
A;Reference number: A9629; MUID:21156231; PMID:11258796
A;Accession: E90589
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-663 <HAY>
A;Cross-references: UNIPROT:P18401; UNIPARC:UPI0000111624; GB:BA000007; PIDN:BA033908.1;
A;Experimental source: strain O157:H7, substrain RMD 050952
C;Genetics:
A;Gene: Ec0485
C;Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology
C;Keywords: copper; electron transfer; heme; iron; magnesium; membrane-associated complex
F;106-421/Binding site: heme a iron (His) (axial ligands) #status predicted
F;284-288/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted
F;288/Binding site: oxygen (Tyr) #status predicted
Query Match 2.9%; Score 101.5; DB 2; Length 663;
Best Local Similarity 22.2%; Pred. No. 6.7;
Matches 63; Conservative 35; Mismatches 95; Indels 91; Gaps 15;
QY 235 IPYVPLRAPPYVWIGRLPSVGHALHEERVPMPPTFLYTQSWEDPDPMEVMEINP 294
Db 179 IEYSPGVGDYVWISLQSLGIGTTLTGIN-----PFVTLKMR-----PGMTMFKM-P 226
QY 295 KDTVLTITSGCNALNLLVQAGQVSDCNPAQSALELKKVAIQOLEPED----- 346
Db 227 VFTWASLCA-----NVLIASFPIITVTV-----ALLTLDYLGTHFTNDMGGNNMMY 275
QY 347 ---VMOLFGEVHPRIEELYEKKLAPF--LSQTSNFWSKRLWYFOHGLYYQGMGKLCW 401
Db 276 INLIW-AWG---HP---EVYILLPVGFSEIAATFSRKLRF-----GYTSLWV 318
QY 402 VLOCLAVV-----LGLKTVKRLANAPTMBEQRRLWDSNMLIHFKVNGPKPLVW 450
Db 319 ATVCITVLSPIVHLHFFTMGAGANVNAFFGITM-----IIAIPGVKIFNW 366
QY 451 LFKVFSVLVFNKAVLM-----FGG-----GVPGKQYAL 479
Db 367 LFTWYQGRIVFHSAMLWTIGFIVTFSVGGMTGVLLAVPGADPVL 410
RESULT 15
B42226
bo-type ubiquinol oxidase (EC 1.10.3.-) chain I - *Escherichia coli* (strain K-12)
N;Alternate names: cytochrome bo chain I
C;Species: *Escherichia coli*
C;Date: 19-Jun-1992 #sequence revision 19-Jun-1992 #text_change 09-Jul-2004
C;Accession: B42226; G64772; JC5635
R;Chapuri, V.; Lemieux, L.; Au, D.C.T.; Gennis, R.B.
J. Biol. Chem. 265, 11185-11192, 1990
A;Title: The sequence of the cyo operon indicates substantial structural similarities be
A;Reference number: A42226; MUID:90293062; PMID:2162835
A;Accession: B42226
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-663 <CHE>
A;Cross-references: UNIPROT:P18401; UNIPARC:UPI0000111624; GB:J05492; NID:gl45651; PIDN:
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of *Escherichia coli* K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: G64772
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-663 <BLAT>
A;Cross-references: UNIPARC:UPI0000111624; GB:AE000149; GB:U00096; NID:gl786628; PIDN:AP
A;Experimental source: strain K-12, substrain MG1655

R;Kawasaki, M.; Mogi, T.; Anraku, Y.
J. Biochem. 122, 422-429, 1997
A;Title: Substitutions of charged amino acid residues conserved in subunit I perturb the
A;Reference number: JC5635; MUID:98021083; PMID:9378723
A;Accession: JC5635
A;Molecule type: protein
A;Residues: 1-55, 'DI', '58-300', 'AI', '303-330', 'LM', '333-510', 'G', '512-663' <KAW>
A;Cross-references: UNIPARC:UPI0000175173
C;Comment: This protein binds all the redox metal centers, low-spin heme b, high-spin heme
C;Genetics:
A;Gene: cyoB
C;Complex: heterooligomer; the cyoABCDE gene products are required for bo-type ubiquinol
ubunits II, III and IV may be required for the assembly of the metal centers in subunit I
C;Function:
A;Description: the cytochrome o complex catalyzes the oxidation of ubiquinol to ubiquinol
C;Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology
C;Keywords: chromoprotein; copper; electron transfer; heme; iron; magnesium; membrane-as
F;16-32/Domain: transmembrane #status predicted <TM1>
F;53-501/Domain: cytochrome-c oxidase chain I homology <CO1>
F;57-73/Domain: transmembrane #status predicted <TM2>
F;107-123/Domain: transmembrane #status predicted <TM3>
F;146-162/Domain: transmembrane #status predicted <TM4>
F;237-253/Domain: transmembrane #status predicted <TM5>
F;287-303/Domain: transmembrane #status predicted <TM6>
F;316-332/Domain: transmembrane #status predicted <TM7>
F;347-363/Domain: transmembrane #status predicted <TM8>
F;387-403/Domain: transmembrane #status predicted <TM9>
F;414-430/Domain: transmembrane #status predicted <TM10>
F;461-477/Domain: transmembrane #status predicted <TM11>
F;498-514/Domain: transmembrane #status predicted <TM12>
F;591-607/Domain: transmembrane #status predicted <TM13>
F;610-626/Domain: transmembrane #status predicted <TM14>
F;106-421/Binding site: heme b iron (His) (axial ligands) #status predicted
F;284-333-334/Binding site: copper (His) #status predicted
F;284-288/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted
F;288/Binding site: oxygen (Tyr) #status predicted
F;419/Binding site: heme o iron (His) (axial ligand) #status predicted
Query Match 2.9%; Score 101.5; DB 2; Length 663;
Best Local Similarity 22.2%; Pred. No. 6.7;
Matches 63; Conservative 35; Mismatches 95; Indels 91; Gaps 15;
QY 235 IPYVPLRAPPYVWIGRLPSVGHALHEERVPMPPTFLYTQSWEDPDPMEVMEINP 294
Db 179 IEYSPGVGDYVWISLQSLGIGTTLTGIN-----PFVTLKMR-----PGMTMFKM-P 226
QY 295 KDTVLTITSGCNALNLLVQAGQVSDCNPAQSALELKKVAIQOLEPED----- 346
Db 227 VFTWASLCA-----NVLIASFPIITVTV-----ALLTLDYLGTHFTNDMGGNNMMY 275
QY 347 ---VMOLFGEVHPRIEELYEKKLAPF--LSQTSNFWSKRLWYFOHGLYYQGMGKLCW 401
Db 276 INLIW-AWG---HP---EVYILLPVGFSEIAATFSRKLRF-----GYTSLWV 318
QY 402 VLOCLAVV-----LGLKTVKRLANAPTMBEQRRLWDSNMLIHFKVNGPKPLVW 450
Db 319 ATVCITVLSPIVHLHFFTMGAGANVNAFFGITM-----IIAIPGVKIFNW 366
QY 451 LFKVFSVLVFNKAVLM-----FGG-----GVPGKQYAL 479
Db 367 LFTWYQGRIVFHSAMLWTIGFIVTFSVGGMTGVLLAVPGADPVL 410
Search completed: March 14, 2006, 01:58:35
Job time : 28 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: March 14, 2006, 01:58:16 ; Search time 31 Seconds
(without alignments)
1728.188 Million cell updates/sec

Title: US-10-620-914-45

Perfect score: 3463

Sequence: 1 MSGRDRPASVTKNFKSLE.....RVNYSFFYMARKGAKDN 648

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgn2_6/prodata/1/iaa/5_COMB.pep.*
- 2: /cgn2_6/prodata/1/iaa/6_COMB.pep.*
- 3: /cgn2_6/prodata/1/iaa/H_COMB.pep.*
- 4: /cgn2_6/prodata/1/iaa/PCTUS_COMB.pep.*
- 5: /cgn2_6/prodata/1/iaa/RE_COMB.pep.*
- 6: /cgn2_6/prodata/1/iaa/backfilese1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	352	10.2	310	2	US-09-248-796A-21233
2	158	4.6	222	2	US-09-248-796A-26335
3	121	3.5	212	2	US-09-107-532A-7254
4	118	3.4	329	2	US-09-902-540-16360
5	111	3.2	195	2	US-09-583-110-4761
6	111	3.2	198	2	US-09-107-433-4622
7	108.5	3.1	706	2	US-09-252-991A-24766
8	107	3.1	255	2	US-09-667-373-2
9	107	3.1	255	2	US-10-305-413-2
10	104.5	3.0	677	2	US-09-489-039A-13088
11	103.5	3.0	1070	2	US-09-107-532A-3716
12	101.5	2.9	663	2	US-09-711-164-332
13	99.5	2.9	471	2	US-08-994-689C-1
14	98.5	2.8	555	2	US-09-134-078-24
15	98.5	2.8	2636	2	US-09-252-991A-25753
16	97.5	2.8	464	2	US-09-134-001C-3488
17	97.5	2.8	471	2	US-08-994-689C-21
18	97	2.8	2209	2	US-10-017-754-1903
19	96.5	2.8	555	2	US-09-252-991A-28313
20	96.5	2.8	2470	2	US-08-265-967C-2
21	96.5	2.8	2470	2	US-08-305-790B-3
22	96	2.8	717	2	US-09-248-796A-18993
23	95.5	2.8	459	2	US-09-949-016-7329
24	94	2.7	480	2	US-09-438-185A-934
25	93.5	2.7	445	2	US-09-252-991A-17629
26	93.5	2.7	471	2	US-09-391-104-25
27	93.5	2.7	475	2	US-09-248-796A-15861

28	93.5	2.7	480	2	US-09-949-016-10560	Sequence 10560, A
29	93.5	2.7	590	2	US-09-489-039A-10743	Sequence 10743, A
30	92	2.7	574	2	US-09-605-703B-254	Sequence 254, App
31	92	2.7	1036	2	US-09-489-039A-10266	Sequence 10266, A
32	91.5	2.6	341	2	US-09-538-092-634	Sequence 634, App
33	91.5	2.6	681	2	US-09-248-796A-16367	Sequence 16367, A
34	91	2.6	749	2	US-10-104-047-2622	Sequence 2622, App
35	90.5	2.6	504	1	US-08-645-900A-1	Sequence 1, Appli
36	90.5	2.6	504	1	US-08-882-238A-1	Sequence 1, Appli
37	90.5	2.6	504	1	US-08-667-790A-1	Sequence 1, Appli
38	90.5	2.6	504	2	US-09-220-459-1	Sequence 1, Appli
39	90.5	2.6	504	2	US-08-938-669A-32	Sequence 32, Appli
40	90.5	2.6	504	2	US-08-546-568B-1	Sequence 1, Appli
41	90.5	2.6	504	2	US-08-822-999-3	Sequence 3, Appli
42	90.5	2.6	504	2	US-09-056-285A-8	Sequence 8, Appli
43	90.5	2.6	504	2	US-09-306-828-32	Sequence 32, Appli
44	90.5	2.6	504	2	US-09-952-464A-8	Sequence 8, Appli
45	90	2.6	500	2	US-09-543-681A-6847	Sequence 6847, Ap

ALIGNMENTS

RESULT 1

US-09-248-796A-21233
; Sequence 21233, Application US/09248796A
; Patent No. 6747137

; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS

; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132

; CURRENT APPLICATION NUMBER: US/09/248,796A

; CURRENT FILING DATE: 1999-02-12

; PRIOR APPLICATION NUMBER: US 60/074,725

; PRIOR FILING DATE: 1998-02-13

; PRIOR APPLICATION NUMBER: US 60/096,409

; PRIOR FILING DATE: 1998-08-13

; NUMBER OF SEQ ID NOS: 28208

; SEQ ID NO 21233

; LENGTH: 310

; TYPE: PRT

; ORGANISM: Candida albicans

US-09-248-796A-21233

Query Match 10.2%; Score 352; DB 2; Length 310;

Best Local Similarity 29.6%; Pred. No. 3 le-30;

Matches 95; Conservative 35; Mismatches 97; Indels 94; Gaps 8;

QY 144 LITFSYSLTMIPTFNNAIDNAVSKLDMEGIIATVDFGIQSSDTSMDGRINTVGGVLRNDIP 62

DB 3 LITFSYSLTMIPTFNNAIDNAVSKLDMEGIIATVDFGIQSSDTSMDGRINTVGGVLRNDIP 62

QY 194 WRRRPFWRIFPIDNIDIGPERRAVLEOKLERWEQNTQ-----GSIPYVPLRAPPYVWI 249

DB 63 WILRNFWRIFWADKFLDSSRRNYLEYKFGTVKSLNSYNKALGKI-----PYIWI 114

QY 250 GRLPSVGHALHEE-----RVERPMPFPT-----273

DB 115 GCDKSKSHTILRLNCLATESPYLATPTTPIANQLEDIPISKHEAALINLQNLVPSM 174

QY 274 -----FLYTOSWEDPEPDMVMEINPKDVTTLTSGGNAL 309

DB 175 YVOKEYRWVYDEMNPVLEQKNQYIYFTWEDPRDHKLNFSTSDTVLAITSAGDNIL 234

QY 310 NL--LVQAGQVVSVDNCPAQSAALLEKVAQOOLEFEDVWOLFGEVHPRIEELK 367

DB 235 SYASLTPPKTKHVDLNPQNHLLKLASFCLSQEQIWSMGEGKIENFNDLIDTL 294

QY 368 APPLSQTSNFWSKRLWYFQH 388

DB 295 APHMSNA-----FQH 305

```
RESULT 2
US-09-248-796A-26335
; Sequence 26335, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 26335
; LENGTH: 222
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-26335

Query Match 4.6%; Score 158; DB 2; Length 222;
Best Local Similarity 35.1%; Pred. No. 9.7e-09;
Matches 34; Conservative 20; Mismatches 19; Indels 24; Gaps 3;

Qy 51 LESFYGQAAAF-----AARLAERSNLIWDLGGGTGCTGVNDMMADYI 92
Db 123 LESFYKQAAHIYDNTREFLLKGRQCLRLAISHLPKKKDLIWDIGGGTGSNIEFMDEIS 182
Qy 93 DLAK-FKSIYVVDLCHSLCEVAKKAKA-----KGWK 123
Db 183 KISENFAVIYDLSPLCEVAKARFESHDLTFRDWK 219

RESULT 3
US-09-107-532A-7254
; Sequence 7254, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 7254:
```

```
SEQUENCE CHARACTERISTICS:
LENGTH: 212 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc_feature
LOCATION: (B) LOCATION 1...212
SEQUENCE DESCRIPTION: SEQ ID NO: 7254:
US-09-107-532A-7254

Query Match 3.5%; Score 121; DB 2; Length 212;
Best Local Similarity 30.8%; Pred. No. 0.00013;
Matches 33; Conservative 23; Mismatches 44; Indels 8; Gaps 3;

Qy 73 IWVDLGGGTGCTGVNDMMADYIDLAKFKSIYVVDLCHSLCEVAKKAKAKGKWKVQVVEADA 132
Db 53 VLLDYGCGTGLVSLPLAE-----RFKELIADASETMLKWAEEKIQADLKNVRTIHADA 107
Qy 133 -COFAPPEGTATLITFSYSLTMIPPHNVVDQACSVLSQDGLGVADF 179
Db 108 SVEF--PAVQANLILLSLVLLHIPDPTENILTKLYEILAPGGQLIIVDF 153

RESULT 4
US-09-902-540-16360
; Sequence 16360, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 16360
; LENGTH: 329
; TYPE: PRT
; ORGANISM: Myxococcus xanthus
US-09-902-540-16360

Query Match 3.4%; Score 118; DB 2; Length 329;
Best Local Similarity 21.8%; Pred. No. 0.00059;
Matches 81; Conservative 37; Mismatches 121; Indels 132; Gaps 16;

Qy 264 VERPPMFPPTLYTQSWEDPEPDMVMEINPKDVTLTGSGGNALMLVQ-GAGQVSV 322
Db 6 VSTPPL---RLKEAVVREDAALELALVERTRARAVLTVASGGCTLLTLARRHPALELVGF 62
Qy 323 DCNPAQSALLELKKVAIQQL-----EFEDVHQLFG-----EGVHPRIEELYEKKLAPFL 371
Db 63 DFNPRQLAHVREKAEGRLPLARYSDAEDAALNQRGEFGLFRTLRRFIEEFVAP-- 120
Qy 372 SQTSHNFWSKRLWYFQGLYYQGMGKLCWLVLCGLVGLGKTVKELANAP-TMBEQRR 430
Db 121 ---AHE-----LAAFF-----APATTASQRR 138
Qy 431 LWDNSMLIHFKNGPKPLVWLFKVLSVLFNKAVLWFGGVGPKQVALLKADGIPENY 490
Db 139 -----EACARWFAS--PYWPVAFELALAAPLN- 164
Qy 491 IARTMDGVAENSHVRKQNY-----FYNCILGTGKFLRDNCPTYL 529
Db 165 ---TMFGPAATQHAEPGSGYPGQVAFERGLQREDAPRNPFLQHVLLGLVLRDAPPEYL 221
```

QY 530 -BAAPA-TLKSUVNLTSTNFMELKARTYTKVILMDHV-DWLDMPVANELAECLAK 586
DB 222 AEGALATLVQSLPDVPRDLRF-----DVLSNIFDWSBDALVAEWAGYLAR 270
QY 587 QVAPGGIVWR 597
DB 271 EAPGCAVLIR 281

RESULT 5

US-09-583-110-4761
; Sequence 4761, Application US/09583110
; Patent No. 6699703
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; FILE REFERENCE: Pneumoniae for Diagnostics and Therapeutics
; CURRENT APPLICATION NUMBER: US/09/583,110
; CURRENT FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 4761
; LENGTH: 195
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-583-110-4761

Query Match 3.2%; Score 111; DB 2; Length 195;
Best Local Similarity 28.0%; Pred. No. 0.0014;
Matches 40; Conservative 28; Mismatches 59; Indels 16; Gaps 6;

QY 46 DHAARLESFVGPO---AAAPAARLAER-----SNLIWVLDGGGTGENVDMADYIDLAK- 96
DB 4 DFNHKAETDSPKNIPLANLVCAVEKQIDILSDKVIDLDFGGGTG---LLA--LPLAKQ 57
QY 97 FKSIVYVDLCHSLCEVAKKAKAGKWKVQVVEADACQFAPPGTATLITFSYSLTWIPP 156
DB 58 AKSVTLVDISEKMLEQARLKVEQQAIKNIQFLEQDLPK-NPLEKEPDCCLAVSRVLHMPD 116
QY 157 FHNVIDOACSYLSQDGLGVADF 179
DB 117 LDAALSFLFHQLKEDGKLIADF 139

RESULT 6

US-09-107-433-4622
; Sequence 4622, Application US/09107433
; Patent No. 6800744
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE FOR DIAGNOSTICS
; THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: <Unknown>
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/107,433
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/ 085131
FILING DATE: May 12, 1998
APPLICATION NUMBER: 60/051553
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-011
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 4622:
SEQUENCE CHARACTERISTICS:
LENGTH: 198 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Streptococcus pneumoniae
FEATURE:
NAME/KEY: misc.feature
LOCATION: (B) LOCATION 1...198
SEQUENCE DESCRIPTION: SEQ ID NO: 4622:
US-09-107-433-4622

Query Match 3.2%; Score 111; DB 2; Length 198;
Best Local Similarity 28.0%; Pred. No. 0.0015;
Matches 40; Conservative 28; Mismatches 59; Indels 16; Gaps 6;

QY 46 DHAARLESFVGPO---AAAPAARLAER-----SNLIWVLDGGGTGENVDMADYIDLAK- 96
DB 7 DFNHKAETDSPKNIPLANLVCAVEKQIDILSDKVIDLDFGGGTG---LLA--LPLAKQ 60
QY 97 FKSIVYVDLCHSLCEVAKKAKAGKWKVQVVEADACQFAPPGTATLITFSYSLTWIPP 156
DB 61 AKSVTLVDISEKMLEQARLKVEQQAIKNIQFLEQDLPK-NPLEKEPDCCLAVSRVLHMPD 119
QY 157 FHNVIDOACSYLSQDGLGVADF 179
DB 120 LDAALSFLFHQLKEDGKLIADF 142

RESULT 7

US-09-252-991A-24766
; Sequence 24766, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24766
; LENGTH: 706
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24766

Query Match 3.1%; Score 108.5; DB 2; Length 706;
Best Local Similarity 19.6%; Pred. No. 0.026;
Matches 94; Conservative 54; Mismatches 170; Indels 161; Gaps 25;

QY 42 KKGDDHAARLESFYGPQAAAFARLAERSN-----LIWVLDGGGTGENVDMADYI 92

Db 174 REGDELLQRCARGGPEIAACAPLAERLQACQSVBERVALM-LAAQTGELLRLVGPRE 231
Qy 93 DLAKFSIYVVDLCHSLCEFAKKAKAGKWNV-QVVEADACQAPPBGT-----A 142
Db 232 HLVRSEYVPV-----EVERPLARFASWYELFPRESGDPTR---HGTFFDVIRRLP 279
Qy 143 TLITFSYSLTWIPFHNV-----IDQACSYL--SODG----- 172
Db 280 QIAAMGFDVLYFFPIHPIGHTRKGRNNSLRAGADPGSPYAGSDGGHEATHPELGDR 339
Qy 173 -----LVGV-----ADFYVSGKYDLP-LRQMP-WSRRFFWR-----SI 203
Db 340 EDFRLLIVA VREGMELADFATQSPDHPWLEHPCW---FAWRPDGSLRYAENPPKY 396
Qy 204 FDTNIDIGERRAYLEOKLERVWEQNTQGSIPYV-----PWLRL-APYYWI-- 249
Db 397 EDIVNVDF-----YAEQALPSLWEALRDVVLGMVEQGVTLFRVDPNPHTKPLPFWEWLIA 450
Qy 250 ---GRLPVSGHALHEERVERPPM-----FPTFLY-----TQSW 280
Db 451 EVRGRHPQV--IFLSEAFTRPAMMARLGKVGFSQSYTYFTWRNDKQELABYFAELNQPPW 508
Qy 281 ED---PEPDMVMEINPKDVTLTGSGCNALMLLVQAG-----QVVSVDNCPAQ 329
Db 509 RDCVRPNFFVTDPINPWFQSRGPGFLTRAALATMGSLGMYSGFELCEAAALPGKE 568
Qy 330 ALLELKKVAIQOEFEDVWOLFQGVHPRIEELYEKK--LAPFLSQTSNFWKRLWYF 386
Db 569 EYLDSEKYQLRPDYQAPGNIVAE--IARLNRIIRRENPAQLQTHLGFQAYNAMNDRIYF 625
RESULT 8
US-09-667-373-2
; Sequence 2, Application US/09667373
; Patent No. 6524840
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Nehls, Michael C.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: Enzyme-Like Proteins and Polynucleotides Encoding the Same
; FILE REFERENCE: LEX-0043-USA
; CURRENT APPLICATION NUMBER: US/09/667,373
; PRIOR FILING DATE: 2000-09-22
; PRIOR FILING DATE: 1999-09-24
; PRIOR FILING DATE: 2000-01-18
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 255
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-667-373-2
Query Match 3.1%; Score 107; DB 2; Length 255;
Best Local Similarity 22.9%; Pred. No. 0.0065;
Matches 53; Conservative 20; Mismatches 68; Indels 90; Gaps 11;
Qy 213 PERR-AYLEOKLERVWEQNTQGSIPYVWMLRAPYYVWIGRLPSVGHALHEERVERPPMFP 271
Db 14 PERNCGYREVE---YWDQRYQGAADSAP-----YDWFQDFSSFRALL----- 52
Qy 272 PTFLYTQSWEDPEPDMVMEINPKDVTLTGSGCN-----ALMLLVQAGQVVSVDNCPA 327
Db 53 -----EP-----ELRPEDRILVL---GCGNSALSVELFLGGFPNVTSDYSSV 92
Qy 328 QSALLELKKVAIQOEFEDVWOLFQGVHPRIEELYEKK--LAPFLSQTSNFWKRLWYF 386
Db 569 EYLDSEKYQLRPDYQAPGNIVAE--IARLNRIIRRENPAQLQTHLGFQAYNAMNDRIYF 625
RESULT 9
US-09-667-373-2
; Sequence 2, Application US/09667373
; Patent No. 6524840
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Nehls, Michael C.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: Enzyme-Like Proteins and Polynucleotides Encoding the Same
; FILE REFERENCE: LEX-0043-USA
; CURRENT APPLICATION NUMBER: US/09/667,373
; PRIOR FILING DATE: 2000-09-22
; PRIOR FILING DATE: 1999-09-24
; PRIOR FILING DATE: 2000-01-18
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 255
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-667-373-2
Query Match 3.1%; Score 107; DB 2; Length 255;
Best Local Similarity 22.9%; Pred. No. 0.0065;
Matches 53; Conservative 20; Mismatches 68; Indels 90; Gaps 11;
Qy 213 PERR-AYLEOKLERVWEQNTQGSIPYVWMLRAPYYVWIGRLPSVGHALHEERVERPPMFP 271
Db 14 PERNCGYREVE---YWDQRYQGAADSAP-----YDWFQDFSSFRALL----- 52
Qy 272 PTFLYTQSWEDPEPDMVMEINPKDVTLTGSGCN-----ALMLLVQAGQVVSVDNCPA 327
Db 53 -----EP-----ELRPEDRILVL---GCGNSALSVELFLGGFPNVTSDYSSV 92
Qy 328 QSALLELKKVAIQOEFEDVWOLFQGVHPRIEELYEKK--LAPFLSQTSNFWKRLWYF 386
Db 569 EYLDSEKYQLRPDYQAPGNIVAE--IARLNRIIRRENPAQLQTHLGFQAYNAMNDRIYF 625

Db 93 VVAAMQARYAHVPQLRWETMDVRKLDPPSPASFDVLEKGTLDALLAGERDPTVTSSEGVH 152
Qy 357 PRIEELYE--KKLAP---FLSQTS-----HNFWSKRLWYFQHGILYYQG 394
Db 153 TVDQVLSVSRVLVPGGRFISMTSAAPHFRTRHYAQAYYGWSLRHATYGS 203
RESULT 9
US-10-305-413-2
; Sequence 2, Application US/10305413
; Patent No. 6835564
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Nehls, Michael C.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: Enzyme-Like Proteins and Polynucleotides Encoding the Same
; FILE REFERENCE: LEX-0043-USA
; CURRENT APPLICATION NUMBER: US/10/305,413
; CURRENT FILING DATE: 2002-11-25
; PRIOR FILING DATE: US/09/667,373
; PRIOR FILING DATE: 2000-09-22
; PRIOR FILING DATE: 1999-09-24
; PRIOR FILING DATE: 2000-01-18
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 255
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-305-413-2
Query Match 3.1%; Score 107; DB 2; Length 255;
Best Local Similarity 22.9%; Pred. No. 0.0065;
Matches 53; Conservative 20; Mismatches 68; Indels 90; Gaps 11;
Qy 213 PERR-AYLEOKLERVWEQNTQGSIPYVWMLRAPYYVWIGRLPSVGHALHEERVERPPMFP 271
Db 14 PERNCGYREVE---YWDQRYQGAADSAP-----YDWFQDFSSFRALL----- 52
Qy 272 PTFLYTQSWEDPEPDMVMEINPKDVTLTGSGCN-----ALMLLVQAGQVVSVDNCPA 327
Db 53 -----EP-----ELRPEDRILVL---GCGNSALSVELFLGGFPNVTSDYSSV 92
Qy 328 QSALLELKKVAIQOEFEDVWOLFQGVHPRIEELYEKK--LAPFLSQTSNFWKRLWYF 386
Db 569 EYLDSEKYQLRPDYQAPGNIVAE--IARLNRIIRRENPAQLQTHLGFQAYNAMNDRIYF 625
RESULT 10
US-09-489-039A-13088
; Sequence 13088, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709,2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342

; SEQ ID NO 13088
; LENGTH: 677
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-13088

Query Match 3.0%; Score 104.5; DB 2; Length 677;
Best Local Similarity 21.4%; Pred. No. 0.068;
Matches 63; Conservative 28; Mismatches 91; Indels 111; Gaps 15;
QY 235 IPYVPLRAPYVWIGRLPSVGHALH-----EERVERPMPFP-PTLYTQSWEDPE 284
DB 193 IEYSPGVVDYVWALQSLGIGTTLTGINFVVIKRAAGMTFMPVF-----SWASLC 248
QY 285 PDMEVMEINPKDVTLTSTGGCNALNLLVQAGQVSDCNPAQSALLEKKAIAOOLEF 344
DB 249 ANILLIATSP---ILVT-----IALLTLDLYLGTHTFT 279
QY 345 ED-----VWQLFGGVHPRIEELYEKKLAPF--LSQTSHPFWSKRLWYFQGLY 391
DB 280 NDGMGNMMYINLIW-AWG---HP---EVYILVLPVGFVSEIATFSKRLP----- 325
QY 392 YQCGMGKLCVLOC LAV-----LGLGKTVKRLANAPTEBQRRLWDSNMLIHF 440
DB 326 ---GYTSLWATVCIIVLSFIWVLLHFFTWGAGANVNAFFGIITM-----IIA 370
QY 441 VKNGKPLVWLFVKFVSLVLFNKAFLW-----FGG-----GVPGKQYAL 479
DB 371 IPTGVKIFNWLFTMYQGRIVFNSAMMTIGFIVTFVSGMTGVLLAVPGADFVL 424

RESULT 11

US-09-107-532A-3716
; Sequence 3716, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 3716:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1070 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...1070
; SEQUENCE DESCRIPTION: SEQ ID NO: 3716:
US-09-107-532A-3716

Query Match 3.0%; Score 103.5; DB 2; Length 1070;
Best Local Similarity 22.2%; Pred. No. 0.2;
Matches 80; Conservative 54; Mismatches 113; Indels 113; Gaps 18;
QY 260 HEERVERPMPFPFTLYTQSWEDPEPDMVMEINPKDVTLTSTGGCNALNLLVQGA--- 316
DB 491 HQPKIE-----VTETGSLDIRFDVTGIERQEIHDVLIQ-----SLLRNDAFYT 533
QY 317 ---GOVSVDCNPAQ--SALLELKKVAIQ-----OLE-----PEDVW 348
DB 534 LENGQVLSFDSSEFQOTSQVLQOLRESIRTEGTIHVPKQGLIIIONLEKSNATSESF 593
QY 349 Q-LFGGVHPRIEELYEKKLAPLSQTSNHF-----WSKRLWYFQGLYQCGMGKLCW 401
DB 594 QTWQDLIHP---ERYQAQLPKGLNATMRDYQKQGRWLKWLGHYQFG----- 638
QY 402 VLQCLAVVLGLGKTVKRLANAPTEBQRRLWDSNMLIHFVKNPKPLW-----LFVK 454
DB 639 --GILADEGLGKTLQTIAPLLSEERKSFSALIV-----APASLIYNWQAEVRKFAP 690
QY 455 FVSLVLFNKAFLWFGGVGKQYALIKADGPIENTARTMDGVAENSHVRKQNYFYINC 514
DB 691 SLSIQVIN-----GNKKEREELAKDTRIVTSYASLRQD--LANYQSQKIDYL---I 738
QY 515 LTGKFLRDNCP-----YLREAA-----FATLKSGVDNL-----TVSTNFEELKAR 558
DB 739 LDEAQVKNSSKTAQALRELAVPQRFALSGTPIENNLLELWLSLFAITMPPFTTKFR 798

RESULT 12

US-09-711-164-332
; Sequence 332, Application US/09711164
; Patent No. 6589738
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE
; FILE REFERENCE: ELITRA.008A
; CURRENT APPLICATION NUMBER: US/09/711,164
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: US 60/164415
; PRIOR FILING DATE: 1999-11-9
; NUMBER OF SEQ ID NOS: 469
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 332
; LENGTH: 663
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-711-164-332

Query Match 2.9%; Score 101.5; DB 2; Length 663;
Best Local Similarity 22.2%; Pred. No. 0.14;
Matches 63; Conservative 35; Mismatches 95; Indels 91; Gaps 15;
QY 235 IPYVPLRAPYVWIGRLPSVGHALHEERVERPMPFPFTLYTQSWEDPEPDMVMEINP 294
DB 179 IEYSPGVVDYVWISLQSLGIGTTLTGIN-----PFVTLKRA-----PGMTMFKM-P 226
QY 295 KDTVLTSTGGCNALNLLVQAGQVSDCNPAQSALLEKKAIAOOLEF----- 346
DB 227 VFTWASLCA-----NVLIIASFPIITVTV-----ALLTLDLYLGTHTFTNDGMNMMY 275
QY 347 ---VMQLFGGVHPRIEELYEKKLAPF--LSQTSHPFWSKRLWYFQGLYQCGMGKLCW 401

Db	276	INLIW-AWG---HP---EVYLLIPVGVFSEIAATFSRKLFP-----GYTSLVM 318	
Qy	402	VLOCLAVW-----LGLGKTVKRLANAPTMEQRRLWDSNMLLIHFVKNKPKPLVW 450	
Db	319	ATVCITVLSPFVWLHFFMTGAGANVAFGIITM-----IIAIPTVGVKIFNW 366	
Qy	451	LFVKFVSLVLFNKAVLW-----FGG-----GVPGKQYAL 479	
Db	367	LFTMYOGRIVFHSAMLWTGIFVTFSVGGMTGVLLAVPGADFVL 410	
RESULT 13			
US-08-994-689C-1			
; Sequence 1, Application US/08994689C			
; Patent No. 6613958			
; GENERAL INFORMATION:			
; APPLICANT: Neuhold, Lisa			
; APPLICANT: Killar, Loran			
; TITLE OF INVENTION: TRANSGENIC ANIMAL MODEL FOR			
; TITLE OF INVENTION: DEGENERATIVE DISEASES OF CARTILAGE			
; NUMBER OF SEQUENCES: 21			
; CORRESPONDENCE ADDRESS:			
; ADDRESSEE: Darby & Darby PC			
; STREET: 805 Third Avenue			
; CITY: New York			
; STATE: NY			
; COUNTRY: USA			
; ZIP: 10022			
; COMPUTER READABLE FORM:			
; MEDIUM TYPE: Diskette			
; COMPUTER: IBM Compatible			
; OPERATING SYSTEM: DOS			
; SOFTWARE: FastSeq for Windows Version 2.0			
; CURRENT APPLICATION DATA:			
; APPLICATION NUMBER: US/08/994,689C			
; FILING DATE: 1997-12-19			
; CLASSIFICATION: 800			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER:			
; FILING DATE:			
; ATTORNEY/AGENT INFORMATION:			
; NAME: Green, Reza			
; REGISTRATION NUMBER: 38,475			
; REFERENCE/DOCKET NUMBER: 0630/0DS32			
; TELEPHONE: 212-527-7700			
; TELEFAX: 212-753-6237			
; TELEX: 236687			
; INFORMATION FOR SEQ ID NO: 1:			
; SEQUENCE CHARACTERISTICS:			
; LENGTH: 471 amino acids			
; TYPE: amino acid			
; STRANDEDNESS: single			
; TOPOLOGY: linear			
; MOLECULE TYPE: protein			
US-08-994-689C-1			
Query Match 2.9%; Score 99.5; DB 2; Length 471;			
Best Local Similarity 18.4%; Pred. No. 0.13;			
Matches 100; Conservative 61; Mismatches 159; Indels 223; Gaps 27;			
Qy	51	LESFYGP-----QAAFAARLAERNLIWDLGGGTGEN-VDM-----ADYDLA 95	
Db	43	LRSYTHPTNLAGILKENAASMTLERMQSFGLVETGKLDNDTLSDVMKKPRGVVDVG 102	
Qy	96	KFKSI-----YVVDLCHSLCEVAKKAKAKAKGNVQVVEADACQAPP 138	
Db	103	EYNVFPRLKWSKNLYRINYVTPDMTHSEVEKAFKA-FKWSDTVPLN----- 152	
Qy	139	EGTATLTFYSYSLTMIPFFHNVIDOACSYLSQDGLGVGVADFY-VSGKYDLPRLQMPWSRR 197	
Db	153	-----FTRLHDGIADIMISFGIKBHGDFYPPDFPGSGLLAHAFPPGPN 194	
Qy	198	FWRSIFDIDNIDIGPERRAYLEQKLERVWEQNTQGISIPYVFWLRAPIYVWIGRLPSVGH 257	
Db	195	YGGDAHFDD-----ETWSSKG-----YNLFVAAHEFGH 256	
Qy	258	ALHEERVERPP--MPPPTFLYTSQWEDPDPDMVEMEI-----NPKOTVLTLSG 304	
Db	227	SLGLDHSKDPGALMF-PIYTYTGKSHFMLPDDDDVOGISOGLYGPGEDEPNPKH----- 277	
Qy	305	GCNALNLLVQAGQVVSVDNCPAQS--ALLELKKVAIQOLEPED--VWOLFSGGVHP--- 357	
Db	278	-----PKTPDKDPSLSLDAITSURG---ETMIFKDRFFWRJ-----HPQOV 316	
Qy	358	-----RIBELYEKKLAPFLSQTSHNFWSKRLWYFQHGLYIOGGMKLCW 401	
Db	317	DAELFLTCKSPWELPNRIIDAAVE-----HPSHDL-----IFIFR-----GRKFW 355	
Qy	402	VLOCLAVW-----LGLGKTVKRLANAPTMEQRRLWDSNMLLIHFVKNKPKPLVWLF 452	
Db	356	ALNGYDILEGYPKKISELGLPKVEVKISAA-----VHFEDTG----- 392	
Qy	453	VKFVSLVLFNKAVLW---FGGVPKQYALIKADGIPNIYIARTWDGVAENSHVRKQNY 509	
Db	393	-----KTLFSGNQWRYDDTNHMDKXPRLIEEDFP---GIGDKVDAYVE-----KNGY 440	
Qy	510	FYV 512	
Db	441	IYF 443	
RESULT 14			
US-09-134-078-24			
; Sequence 24, Application US/09134078			
; Patent No. 636844			
; GENERAL INFORMATION:			
; APPLICANT: Bylina, Edward J.			
; TITLE OF INVENTION: GLYCOSIDASE ENZYMES			
; NUMBER OF SEQUENCES: 72			
; CORRESPONDENCE ADDRESS:			
; ADDRESSEE: Gray Cary Ware & Freidenrich LLP			
; STREET: 4365 Executive Drive, Suite 1600			
; CITY: San Diego			
; STATE: CA			
; COUNTRY: USA			
; ZIP: 92121			
; COMPUTER READABLE FORM:			
; MEDIUM TYPE: Diskette			
; COMPUTER: IBM Compatible			
; OPERATING SYSTEM: Windows95			
; SOFTWARE: FastSeq for Windows Version 2.0			
; CURRENT APPLICATION DATA:			
; APPLICATION NUMBER: US/09/134,078			
; FILING DATE: 13-AUG-1998			
; CLASSIFICATION: 435			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER: 08/949,026			
; FILING DATE: 10-OCT-1997			
; APPLICATION NUMBER: 60/056,916			
; FILING DATE: 06-DEC-1996			
; ATTORNEY/AGENT INFORMATION:			
; NAME: Haile, Lisa A.			
; REGISTRATION NUMBER: 38,347			
; REFERENCE/DOCKET NUMBER: 09010/024002			
; TELECOMMUNICATION INFORMATION:			
; TELEPHONE: 858/677-1456			
; TELEFAX: 858/677-1465			
; INFORMATION FOR SEQ ID NO: 24:			
; SEQUENCE CHARACTERISTICS:			
; LENGTH: 555 amino acids			
; TYPE: amino acid			
; TOPOLOGY: linear			
; MOLECULE TYPE: protein			
; FRAGMENT TYPE: internal			

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: March 14, 2006, 01:59:26 ; Search time 15 Seconds
(without alignments)
1202.467 Million cell updates/sec

Title: US-10-620-914-45
Perfect score: 3463
Sequence: 1 MGSGRDGRPASVTKNKNFSLB.....RVNMYSSFYMARCKGAKXDN 648

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 161667 seqs, 27834885 residues

Total number of hits satisfying chosen parameters: 161667

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_AA_New.*

- 1: /cgn2_6/prodata/1/pubpaa/US08_NEW_PUB.pep.*
- 2: /cgn2_6/prodata/1/pubpaa/US06_NEW_PUB.pep.*
- 3: /cgn2_6/prodata/1/pubpaa/US07_NEW_PUB.pep.*
- 4: /cgn2_6/prodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 5: /cgn2_6/prodata/1/pubpaa/US09_NEW_PUB.pep.*
- 6: /cgn2_6/prodata/1/pubpaa/US10_NEW_PUB.pep.*
- 7: /cgn2_6/prodata/1/pubpaa/US11_NEW_PUB.pep.*
- 8: /cgn2_6/prodata/1/pubpaa/US60_NEW_PUB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	106	3.1	472	7	US-11-169-041-156
2	102	2.9	448	7	US-11-124-367A-282
3	96	2.8	326	6	US-10-467-657-7224
4	95.5	2.8	2890	7	US-11-115-639-31
5	95.5	2.8	2890	7	US-11-115-639-32
6	95.5	2.8	2890	7	US-11-115-639-33
7	94	2.7	658	7	US-11-096-568A-2582
8	94	2.7	680	7	US-11-096-568A-2581
9	94	2.7	680	7	US-11-096-568A-2580
10	93.5	2.7	1197	6	US-10-055-877-8
11	93.5	2.7	1247	6	US-10-055-877-10
12	93	2.7	417	7	US-11-098-686-11330
13	92.5	2.7	1590	6	US-10-055-877-146
14	92	2.7	316	7	US-11-096-568A-1265
15	91.5	2.6	241	7	US-11-098-686-10262
16	91	2.6	749	7	US-11-072-512-2622
17	89	2.6	537	7	US-11-109-156-28
18	89	2.6	577	6	US-10-718-264-3
19	89	2.6	577	6	US-10-718-264-3
20	88.5	2.6	329	7	US-11-234-786-376
21	87.5	2.5	834	7	US-11-087-099-10005
22	87	2.5	265	6	US-10-467-657-6590
23	87	2.5	265	6	US-10-467-657-7654
24	87	2.5	565	7	US-11-072-512-2180
25	86.5	2.5	395	6	US-10-793-626-664

ALIGNMENTS

RESULT 1

US-11-169-041-156
; Sequence 156, Application US/11169041
; Publication No. US20060019284A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: IDENTIFICATION OF POLYNUCLEOTIDES FOR PREDICTING ACTIVITY OF
; TITLE OF INVENTION: COMPOUNDS THAT INTERACT WITH AND/OR MODULATE PROTEIN TYROSINE
; TITLE OF INVENTION: KINASES AND/OR PROTEIN TYROSINE KINASE PATHWAYS IN LUNG CANCER
; TITLE OF INVENTION: CELLS
; FILE REFERENCE: 10001 NP
; CURRENT APPLICATION NUMBER: US/11/169,041
; PRIOR FILING DATE: 2005-06-28
; PRIOR APPLICATION NUMBER: 60/584,405
; PRIOR FILING DATE: 2004-06-30
; NUMBER OF SEQ ID NOS: 527
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 156
; LENGTH: 472
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-169-041-156

Query Match 3.1%; Score 106; DB 7; Length 472;
Best Local Similarity 20.4%; Pred. No. 0.13;
Matches 90; Conservative 54; Mismatches 155; Indels 142; Gaps 19;

QY	12	YTKQNFLEKLSNMKDDLTVLRHMWFGSKKGDHAA-----RLRSFYGPQAA--AFNA 64
DB	39	FYRTFQNFREFK-ATMNCNLLAVLKHLL-----KGNEAALECLRKABELTQEHADQAEIR 92
QY	65	RLAERSNLWDLGGGTGENVDMDADYID--LAKFKSIYVVDICHSCEVKKAKAGW 122
DB	93	SLVTWGNVAVVYHMGRLSDVQIYVDVKVHCKEFPYRISPELDCBEGWTRKCGN 152
QY	123	KN--VOVVEADACQAP--PEGTATLITFSYSLTMTIPPHNVIDQACSYLSQDGLGVAD 178
DB	153	QNERAKVCPEKALEKKPKNPFTSGLAIASYRLDNWPPSQNAID-----196
QY	179	FVVSQKYDPLQMPWRSRFFWRSIFDIDNIDIGPER---RAYLEOKLRVWQNTQSSI 235
DB	197	-----PLRQA-----IRLNPONQYKLVLLALKHMRSE-----225
QY	236	PYVPMLRAPPYVWIGRLPSVGHALHEERVERPP-----MFPPTFLYTQSWEDPFDMEV-- 289
DB	226	-----GEEGEKEKLVVEALEKAPGTVDLRSAAKFFYRKDEPDKALIELLK 271
QY	290	--MEINPKDVTILTTSGGCNALNLLVQAGQVVSVDNCP--AQSALLELKVAIQOL--- 342

Db 272 KALEYIPNNAYLHCQIGCCYRAKVF-----QVNNLRENGMYGKRKLELIGHAVAHKKA 326
Qy 343 -----BFDVQVLFGBGVHPRIEYKELAPLSQTSF----- 376
Db 327 DEANDNLFRVCSILASHALADQYEDAEYVF-----QKEFSKELTPVAKQLLHRYG 378
Qy 377 NFW-----SKRLWYFQHGL 390
Db 379 NFOLYQMKCEDKAIHHFIEGV 399
RESULT 2
US-11-124-367A-282
; Sequence 282, Application US/11124367A
; Publication No. US20060024700A1
; GENERAL INFORMATION:
; APPLICANT: Michele Cargill
; APPLICANT: Hongjin Huang
; TITLE OF INVENTION: Genetic Polymorphisms Associated with
; TITLE OF INVENTION: Fibrosis Methods of Detection and Uses Thereof
; FILE REFERENCE: CU001519.ORD
; CURRENT APPLICATION NUMBER: US/11/124,367A
; CURRENT FILING DATE: 2005-05-09
; PRIOR APPLICATION NUMBER: US 60/568,846
; PRIOR FILING DATE: 2004-05-07
; PRIOR APPLICATION NUMBER: US 60/582,609
; PRIOR FILING DATE: 2004-06-25
; PRIOR APPLICATION NUMBER: US 60/599,554
; PRIOR FILING DATE: 2004-08-09
; NUMBER OF SEQ ID NOS: 34460
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 282
; LENGTH: 448
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-124-367A-282

Query Match 2.9%; Score 102; DB 7; Length 448;
Best Local Similarity 20.8%; Pred. No. 0.29;
Matches 90; Conservative 54; Mismatches 163; Indels 126; Gaps 19;
Qy 12 YTKKNSLEKLSMKDDTLVLRHWMFGSKGDDHAA-----RLSEFYQQA--AFAA 64
Db 15 FYTEFQNRFEK-ATMCNLLAYLKLH-----KQNEAALECLKAEELIQEHADQAEIR 68
Qy 65 RLAEBSNLIWVJGGGTGENVDMADYID--LAKFSIYVVDLCHSLCEVAKKKAKAGW 122
Db 69 SLVTWGNVAVVYHMGRLSDVQIYVDKVHVCBKFSPYRIESPELDCBEGWRLKCGGN 128
Qy 123 KN--VQVVEADACQAP--PEGTATLITFSYSLTMIPPFHNVIDQACSYLSQDGLGVAD 178
Db 129 QNERAKVCFBKALEKPKNPEFTSGLAIASYRLDNWPPSQNAID----- 172
Qy 179 FYVSGKYDLPLRQMPWRRFRFWSIFDIDNIDIGPER---RAYLEOKLERVWEQNTQGS 235
Db 173 -----PIRQA-----IRLNPQYKLVLLALKHKRRE----- 201
Qy 236 PYVPWLRAPYVWIGRLPSVGHALHEERVRPP-----MFPPTFLYTQSWEDPDPMEV-- 289
Db 202 -----GEBEGEKLVEALEKAPGVTDVLSAAKFYRRKDEPKAIELLK 247
Qy 290 --MEINPKDVTILTSGGCNALNLLVQAGQVVSVCNP--AQSALELAKVAIQOLEFE 345
Db 248 KALEYIPNNAYLHCQIGCCYRAKVF-----QVNNLRENGMYGKRKLELIGHAVAHKKA 302
Qy 346 D-----VWQLFG-----EGVHPRIEELYEKKLAPFLSQTSF-----NFW----- 379
Db 303 DEANDNLFRVCSILASHALADQYEEAEYVFQKEFSKELTPVAKQLLHRYGNFQYQMK 362
Qy 380 --SKRLWYFQHGL 390
Db 363 CEDKAIHHFIEGV 375

RESULT 3
US-10-467-657-7224
; Sequence 7224, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SPA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 7224
; LENGTH: 326
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-7224

Query Match 2.8%; Score 96; DB 6; Length 326;
Best Local Similarity 25.9%; Pred. No. 0.64;
Matches 30; Conservative 21; Mismatches 39; Indels 26; Gaps 6;
Qy 120 KGKMNQVVEADACQAPP---EGTATLITFSYSLTMIPPF-----HNV--IDOACS--- 166
Db 214 EGWLEHEVTRLTVCEYLKPLADGIDTLVLGCTHFFLLKPLIGREAHNVALVDSAITTAE 273
Qy 167 ----YLSQDLGVAD-----FYSGKYDLPLRQMPWRRFRFWSIFDIDNIDIG 212
Db 274 ETARVLAQEGLLDTGNNPDYRFYVS---DIPLKFRITIGERFLGRTMEQIEMVSLG 326

RESULT 4
US-11-115-639-31
; Sequence 31, Application US/11115639
; Publication No. US2005028242A1
; GENERAL INFORMATION:
; APPLICANT: Rothstein, David
; APPLICANT: Murphy, Christopher
; APPLICANT: MacNeil, Ian
; TITLE OF INVENTION: SCREENING ASSAYS FOR ANTIMICROBIAL
; FILE REFERENCE: 50150/075003
; CURRENT APPLICATION NUMBER: US/11/115,639
; CURRENT FILING DATE: 2005-04-27
; PRIOR APPLICATION NUMBER: 60/566,858
; PRIOR FILING DATE: 2004-04-30
; PRIOR APPLICATION NUMBER: 60/565,679
; PRIOR FILING DATE: 2004-04-27
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 2890
; TYPE: PRT
; ORGANISM: H. pylori
US-11-115-639-31

Query Match 2.8%; Score 95.5; DB 7; Length 2890;
Best Local Similarity 18.9%; Pred. No. 17;
Matches 101; Conservative 82; Mismatches 150; Indels 201; Gaps 29;
Qy 214 ERRAY-----LEOKLERVWE---QNTQGSIPYVPWLRAPYVWIGRLPSVGHALH 260
Db 1775 EERGYATTLKQAKRMIEQKSNVEWECLOEITEG-----YFVLLNRAPI-----LH 1819
Qy 261 EERVERPPMPPTFLYTQSWEDPEPDMVMEINP----- 294

Db 1820 KOSIQ-----AFHPKLI-----DKAIQLHPLVCSAFNADFDGDMVHVPLSQEAI 1866
Qy 295 -KDTVLTLTSGGCNALNLVQAGVSVDCNPAQSALLEKKVAIQLEFEDVWOLFGE 353
Db 1867 AECKVLMSS-----WNILLPASGKAVAI---PSQDWLVGLYLSLEKSGVGEHKLFS 1918
Qy 354 -----GVHPRIEELYE-----KKLAPLSOTSNFWSKRLWYFQ 387
Db 1919 VNEIITAITDKELDIHAKIRVLDQGNIIATISAGRMIIKSILPDFIPT--DLWNRPMKKKD 1976
Qy 388 HG-----LYYQGMGKLCWVLOCL-----AVVLGLG-----KTVKRLANAPTWE 426
Db 1977 IGVLVDYVHKVGIGITATFLDNLKTGLFRYATKAGISISMEDIIITPKDKQKMKVEKAKVE 2036
Qy 427 EQR--RLWDSNMLIHFRVQKPKPLVWLFV-----KFVSLVL-----FNKAVLMFGGGVP 473
Db 2037 VKKIQQYDQGLLTDOERYNKIIDIWTEVNDKMSKEMMTAIAQDKGEGFNSIYMMADSGAR 2096
Qy 474 GK-----QYALIKADGPIENYIARTMDGVAENSHVRKQNYFYNCITG--KFLR 521
Db 2097 GSAAIQIRQLSAMRGLMTKPDGSIITETPI---ISNFKEGNLVLE---YFNSTHGARKGLA 2149
Qy 522 DNCPTVLRFAAFATLAKSGVVDNLTVSTNF-----FMEELKAR 558
Db 2150 D---TALKTANAGYLTRKLIID---VSNQVNVSDDCGTHEGIEITDIAGVSELIEPLEER 2203
Qy 559 TYTKVILMDHVDLMDPANEL---AECL-----AKQVAPGGGIVWRSASLSPP 604
Db 2204 IFRVLELVEDVID---PITNEILLYADTLIDEGAKKVVVEAGI---KSITIRTP 2250

RESULT 5

US-11-115-639-32
; Sequence 32, Application US/11115639
; Publication No. US20050282424I
; GENERAL INFORMATION:
; APPLICANT: Rothstein, David
; APPLICANT: Murphy, Christopher
; APPLICANT: MacNeil, Ian
; TITLE OF INVENTION: SCREENING ASSAYS FOR ANTIMICROBIAL
; FILE OF INVENTION: AGENTS
; FILE REFERENCE: 50150/075003
; CURRENT APPLICATION NUMBER: US/11/115,639
; CURRENT FILING DATE: 2005-04-27
; PRIOR APPLICATION NUMBER: 60/566,858
; PRIOR FILING DATE: 2004-04-30
; PRIOR APPLICATION NUMBER: 60/565,679
; PRIOR FILING DATE: 2004-04-27
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 32
; LENGTH: 2890
; TYPE: PRT
; ORGANISM: H. pylori
US-11-115-639-32

Query Match 2.8%; Score 95.5; DB 7; Length 2890;
Best Local Similarity 18.9%; Pred. No. 17;
Matches 101; Conservative 82; Mismatches 150; Indels 201; Gaps 29;
Qy 214 ERRAY-----LEOKLERWE---QNTQGSIPYVWLRAPYVYVWITGRPLPSVGHALH 260
Db 1775 EERGVAATLTKQAKRMIEQKSNVEWECLOEITEG-----YFVLNRAPT---LH 1819
Qy 261 EERVERPPMPPTFLYTSQWEDPEPDMVEINP----- 294
Db 1820 KOSIQ---AFHPKLI-----DKAIQLHPLVCSAFNADFDGDMVHVPLSQEAI 1866
Qy 295 -KDTVLTLTSGGCNALNLVQAGVSVDCNPAQSALLEKKVAIQLEFEDVWOLFGE 353
Db 1867 AECKVLMSS-----WNILLPASGKAVAI---PSQDWLVGLYLSLEKSGVGEHKLFS 1918

Qy 354 -----GVHPRIEELYE-----KKLAPLSOTSNFWSKRLWYFQ 387
Db 1919 VNEIITAITDKELDIHAKIRVLDQGNIIATISAGRMIIKSILPDFIPT--DLWNRPMKKKD 1976
Qy 388 HG-----LYYQGMGKLCWVLOCL-----AVVLGLG-----KTVKRLANAPTWE 426
Db 1977 IGVLVDYVHKVGIGITATFLDNLKTGLFRYATKAGISISMEDIIITPKDKQKMKVEKAKVE 2036
Qy 427 EQR--RLWDSNMLIHFRVQKPKPLVWLFV-----KFVSLVL-----FNKAVLMFGGGVP 473
Db 2037 VKKIQQYDQGLLTDOERYNKIIDIWTEVNDKMSKEMMTAIAQDKGEGFNSIYMMADSGAR 2096
Qy 474 GK-----QYALIKADGPIENYIARTMDGVAENSHVRKQNYFYNCITG--KFLR 521
Db 2097 GSAAIQIRQLSAMRGLMTKPDGSIITETPI---ISNFKEGNLVLE---YFNSTHGARKGLA 2149
Qy 522 DNCPTVLRFAAFATLAKSGVVDNLTVSTNF-----FMEELKAR 558
Db 2150 D---TALKTANAGYLTRKLIID---VSNQVNVSDDCGTHEGIEITDIAGVSELIEPLEER 2203
Qy 559 TYTKVILMDHVDLMDPANEL---AECL-----AKQVAPGGGIVWRSASLSPP 604
Db 2204 IFRVLELVEDVID---PITNEILLYADTLIDEGAKKVVVEAGI---KSITIRTP 2250

RESULT 6

US-11-115-639-33
; Sequence 33, Application US/11115639
; Publication No. US20050282424I
; GENERAL INFORMATION:
; APPLICANT: Rothstein, David
; APPLICANT: Murphy, Christopher
; APPLICANT: MacNeil, Ian
; TITLE OF INVENTION: SCREENING ASSAYS FOR ANTIMICROBIAL
; FILE OF INVENTION: AGENTS
; FILE REFERENCE: 50150/075003
; CURRENT APPLICATION NUMBER: US/11/115,639
; CURRENT FILING DATE: 2005-04-27
; PRIOR APPLICATION NUMBER: 60/566,858
; PRIOR FILING DATE: 2004-04-30
; PRIOR APPLICATION NUMBER: 60/565,679
; PRIOR FILING DATE: 2004-04-27
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 33
; LENGTH: 2890
; TYPE: PRT
; ORGANISM: H. pylori
US-11-115-639-33

Query Match 2.8%; Score 95.5; DB 7; Length 2890;
Best Local Similarity 18.9%; Pred. No. 17;
Matches 101; Conservative 82; Mismatches 150; Indels 201; Gaps 29;
Qy 214 ERRAY-----LEOKLERWE---QNTQGSIPYVWLRAPYVYVWITGRPLPSVGHALH 260
Db 1775 EERGVAATLTKQAKRMIEQKSNVEWECLOEITEG-----YFVLNRAPT---LH 1819
Qy 261 EERVERPPMPPTFLYTSQWEDPEPDMVEINP----- 294
Db 1820 KOSIQ---AFHPKLI-----DKAIQLHPLVCSAFNADFDGDMVHVPLSQEAI 1866
Qy 295 -KDTVLTLTSGGCNALNLVQAGVSVDCNPAQSALLEKKVAIQLEFEDVWOLFGE 353
Db 1867 AECKVLMSS-----WNILLPASGKAVAI---PSQDWLVGLYLSLEKSGVGEHKLFS 1918
Qy 354 -----GVHPRIEELYE-----KKLAPLSOTSNFWSKRLWYFQ 387
Db 1919 VNEIITAITDKELDIHAKIRVLDQGNIIATISAGRMIIKSILPDFIPT--DLWNRPMKKKD 1976
Qy 388 HG-----LYYQGMGKLCWVLOCL-----AVVLGLG-----KTVKRLANAPTWE 426
Db 1977 IGVLVDYVHKVGIGITATFLDNLKTGLFRYATKAGISISMEDIIITPKDKQKMKVEKAKVE 2036

QY 427 EQR--RLWDSNMLIHFKVNGPKPLVWLFV-----KFSVLVL-----FNKAVLWFGGVP 473
Db 2037 VKKIQOYDGLLTDQERYNKIITDTEVNDKMSKEMMTAIAQDKGFGNSIYMWADSGAR 2096
QY 474 GK-----QYALIKADGIPNIYARTMDGVAENSHVRKQNTFYNYNCLTG--KFLR 521
Db 2097 GSAAQIRQLSAMRGLMTKPDGSIETPI---ISNFKEGNLVLE---YFNSTHGARKGLA 2149
QY 522 DNCPTVILREAFATLKSQVVDNUTVSTNF-----PMSELKAR 558
Db 2150 D---TALKTANAGYLTKLID--VSQNVKVSDDCGTHEGIBITDIAVGSSELIBLEER 2203
QY 559 TYTKVILMDHVDLMPVANEL---AECL-----AKQVAPGGIVIVRSASLSPP 604
Db 2204 IFGRVLLEDVID----PITNEILLYADTLIDBEGAKKVVVETAGI---KSIITIRP 2250

RESULT 7

US-11-096-568A-2582
; Sequence 2582, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096, 568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 2582
; LENGTH: 658
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(658)
; OTHER INFORMATION: Ceres Seq. ID no. 15176030
US-11-096-568A-2582

Query Match 2.7%; Score 94; DB 7; Length 658;
Best Local Similarity 19.6%; Pred. No. 2.7;
Matches 115; Conservative 69; Mismatches 182; Indels 220; Gaps 30;
QY 20 EKLKLSMKDDLTVLRRHMFSGKKGD-----DHAARLESFYGPQAAFAALRAER 69
Db 56 EKLQLAGVSQRCVVL-----GSKAEDKHLQLVHTKDHNVLKSIATKQDYRNRIASQ 110
QY 70 SNLIWDLGGGTGENDVMADYIDLAKFKSIYVVDLCHSLCEVAKKAKAKGKNVQVVE 129
Db 111 LNSIY--LNGSSEAYLAAG-----SVVKLAEKVAEG-----E 142
QY 130 ADACQAPPEGTATLITFSYSLTMIPPFHNVIDOACSYLSQDGLGVADFYVSGKVDLPL 189
Db 143 LD--CGFA-----IVRPPGHAAEADNMGFCLFNNVAVAAASFLNRPDLGV 187
QY 190 RQ--MPW-----SRRPFWR-----IFDIDNIDIGPERRAYLEOKLERVWEQNTQGI 235
Db 188 KKLIVDMDVHNGTGQKMFWDKPRVLFFSVHRHEYGGFYPDGDDGDDYNNVGGSGEGFN 247
QY 236 PYVPWLRAPIYVWIGRLPSVGHALHEERVERPPM--FPPTFLYTSQWEDPEPD----- 286
Db 248 INVPWDQ-----GRCDADYLAADHILIPVAREFNPDIIVLSAGFADAAINDPLGGCC 300
QY 287 -----MEVWEINPKDVTLTSTGGCNALNLVQAGQVVSVDNCPAQSALELKK 336
Db 301 VTPYGYVMLKKLMFPAQGGKIVLAL--EGGYN--LDSTAKSSLACVQV-----LLEDKQ 350
QY 337 V--AIQOLEFEDVWQLFGEVGHPI-----EEL-----YEKKLAPL-----SOTS 375
Db 351 IQGPPAYPESTWRVI--QAVRKELCTYWPSLADELSWKLINQKTPPTPIILISSDSETE 409
QY 376 HN-----FWSKRL-----WYFQHG-----L 390

Db 410 DNAQGLLDQMSKLSIENPOGTLLENHQVEPASTSWRADLAKVDVWYASFGSNMWPFLC 469
QY 391 YYQGGMGKLCWVLOCLAVVLGLKTVKRLANAPTMEEQRELWDSNMLIHFKVNGPKPLVW 450
Db 470 YIQGGQ-----VDGLKKVC-----VGSMD-----KSPPKETVW 497
QY 451 LFWKFSVSLVFNK--AVLWFGGVPKQYALIKADGIPNIYART 494
Db 498 --ETFPRLHLPFGRESSVGVGV-----AFTNPLANLIDQT 532
RESULT 8
US-11-096-568A-2581
; Sequence 2581, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096, 568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 2581
; LENGTH: 660
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(660)
; OTHER INFORMATION: Ceres Seq. ID no. 15176029
US-11-096-568A-2581

Query Match 2.7%; Score 94; DB 7; Length 660;
Best Local Similarity 19.6%; Pred. No. 2.7;
Matches 115; Conservative 69; Mismatches 182; Indels 220; Gaps 30;
QY 20 EKLKLSMKDDLTVLRRHMFSGKKGD-----DHAARLESFYGPQAAFAALRAER 69
Db 58 EKLQLAGVSQRCVVL-----GSKAEDKHLQLVHTKDHNVLKSIATKQDYRNRIASQ 112
QY 70 SNLIWDLGGGTGENDVMADYIDLAKFKSIYVVDLCHSLCEVAKKAKAKGKNVQVVE 129
Db 113 LNSIY--LNGSSEAYLAAG-----SVVKLAEKVAEG-----E 144
QY 130 ADACQAPPEGTATLITFSYSLTMIPPFHNVIDOACSYLSQDGLGVADFYVSGKVDLPL 189
Db 145 LD--CGFA-----IVRPPGHAAEADNMGFCLFNNVAVAAASFLNRPDLGV 189
QY 190 RQ--MPW-----SRRPFWR-----IFDIDNIDIGPERRAYLEOKLERVWEQNTQGI 235
Db 190 KKLIVDMDVHNGTGQKMFWDKPRVLFFSVHRHEYGGFYPDGDDGDDYNNVGGSGEGFN 249
QY 236 PYVPWLRAPIYVWIGRLPSVGHALHEERVERPPM--FPPTFLYTSQWEDPEPD----- 286
Db 250 INVPWDQ-----GRCDADYLAADHILIPVAREFNPDIIVLSAGFADAAINDPLGGCC 302
QY 287 -----MEVWEINPKDVTLTSTGGCNALNLVQAGQVVSVDNCPAQSALELKK 336
Db 303 VTPYGYVMLKKLMFPAQGGKIVLAL--EGGYN--LDSTAKSSLACVQV-----LLEDKQ 352
QY 337 V--AIQOLEFEDVWQLFGEVGHPI-----EEL-----YEKKLAPL-----SOTS 375
Db 353 IQGPPAYPESTWRVI--QAVRKELCTYWPSLADELSWKLINQKTPPTPIILISSDSETE 411
QY 376 HN-----FWSKRL-----WYFQHG-----L 390
Db 412 DNAQGLLDQMSKLSIENPOGTLLENHQVEPASTSWRADLAKVDVWYASFGSNMWPFLC 471
QY 391 YYQGGMGKLCWVLOCLAVVLGLKTVKRLANAPTMEEQRELWDSNMLIHFKVNGPKPLVW 450
Db 472 YIQGGQ-----VDGLKKVC-----VGSMD-----KSPPKETVW 499

QY 451 LFVKFVSLVLFNK--AVLWFGGVGPGKQYALIKADGIPNIYART 494
Db 500 --ETFPRLFFGREGSVGNGGV-----AFTNPLANLIDQT 534

RESULT 9
US-11-096-568A-2580
; Sequence 2580, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096, 568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 2580
; LENGTH: 680
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; NAME/KEY: misc feature
; LOCATION: (1)..(680)
; OTHER INFORMATION: Ceres Seq. ID no. 15176028
US-11-096-568A-2580

Query Match 2.7%; Score 94; DB 7; Length 680;
Best Local Similarity 19.6%; Pred. No. 2.8;
Matches 115; Conservative 69; Mismatches 182; Indels 220; Gaps 30;

QY 20 EKLKLSMKDDLTVLRRHMFSGKGD-----DHAARLESFYGPQAAFAARLAER 69
Db 78 EKLQAGVSQRCVVL-----GSSKAEDKHLQLVHTKDHVNLVKSISTKQDYERNIASQ 132

QY 70 SNLIWDLGGCTGENDVMDADYIDLAKFSIYVDLCHSLCEFAKKAQKAKGWNQVVE 129
Db 133 LNSIY--LNGSGSEAYLAAG-----SYVLAERVAEG-----E 164

QY 130 ADACQAPPEGTATLTFSVSLTMIPPEHNVIDQACSYLSQDGLGVADPYVSGKYDLPL 189
Db 165 LD--CGFA-----IVRPPGHAEADENMGFLFNVAVAASFLMLNERPDLGV 209

QY 190 RQ--MPW-----SRRFFWRS---IFDIDNIDIGPERRAYLEQKLERVWEQNTQSSI 235
Db 210 KKTLIVDMDVHNGCTQKMPKDPRLVFTSVHRHEYGFPAGDDGDDYNNVGGSGEGFN 269

QY 236 PYPVWLRAPYYWIGRLPSVGHALHEERVERPPM--PPPTFLYTQSWDEPPD----- 286
Db 270 INVPMQD-----GRCGDADYLAADHILIPVAREBFNPDVIFLSAGFDAAINDPILGGCC 322

QY 287 -----MEVWEINPKDTVLTLTSGGNCNALNLVQAGQVVSVDNCPAQSALLLEKK 336
Db 323 VTPYGSVMLKKLMFAQKQKIVLAL--EGGYN--LDSIAKSSLACVQV-----LLEDKQ 372

QY 337 V--AIQOLEPQVWQFGEVHPRI-----EEL-----YBKLLAPFL-----SQTS 375
Db 373 IQGPPAYPESTWRVI-QAVRKRLCTYWPSLADELSWKLINQKTPPIILISSDSETE 431

QY 376 HN-----FWSKRL-----WYFQHG-----L 390
Db 432 DNAQGLLDQMSKLSIENPQGLLENHVEPASTSWRADLAKVDVWVASFGSNMKNKPRFLC 491

QY 391 YYQGGWGLCWVLCQAVLVGLCTYVKRLANAPTEEQRLWDSNMLIHPVKNKPKPLW 450
Db 492 YIQGGQ-----VDGLKKVC-----VGSMD-----KSPPKETVM 519

QY 451 LFVKFVSLVLFNK--AVLWFGGVGPGKQYALIKADGIPNIYART 494
Db 520 --ETFPRLFFGREGSVGNGGV-----AFTNPLANLIDQT 554

RESULT 10
US-10-055-877-8
; Sequence 8, Application US/10055877
; Publication No. US20050288241A1
; GENERAL INFORMATION:
; APPLICANT: DeCristofaro, Marc
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Miller, Charles
; APPLICANT: Tchernev, Velizar
; APPLICANT: Zhong, Mei
; APPLICANT: Anderson, David
; APPLICANT: Ballinger, Robert
; APPLICANT: Gerlach, Valerie
; APPLICANT: Spytek, Kimberly
; APPLICANT: Ratelli, Luca
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Guo, Xiaojia
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Andrew, David
; APPLICANT: Mezes, Peter
; APPLICANT: Patturajan, Meera
; APPLICANT: Burgess, Catherine
; APPLICANT: Eissen, Andrew
; APPLICANT: Wolenc, Adam
; APPLICANT: Baumgartner, Jason
; APPLICANT: Shimkets, Richard
; APPLICANT: Gusev, Vladimir
; APPLICANT: Vernet, Corine
; APPLICANT: Taupier Jr., Raymond
; APPLICANT: Pena, Carol
; APPLICANT: Shenoy, Suresh
; APPLICANT: Li, Li
; APPLICANT: Casman, Stacie
; APPLICANT: Boldog, Ference
; TITLE OF INVENTION: Novel Polypeptides and Nucleic Acids Encoded Thereby
; FILE REFERENCE: 21402-251
; CURRENT APPLICATION NUMBER: US/10/055,877
; CURRENT FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: 60/262,892
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 60/263,598
; PRIOR FILING DATE: 2001-01-23
; PRIOR APPLICATION NUMBER: 60/263,799
; PRIOR FILING DATE: 2001-01-24
; PRIOR APPLICATION NUMBER: 60/264,117
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: 60/264,139
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: 60/264,478
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 60/263,351
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/272,870
; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: 60/275,990
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/275,927
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 512
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 8
; LENGTH: 1197
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-055-877-8

Query Match 2.7%; Score 93.5; DB 6; Length 1197;
Best Local Similarity 21.3%; Pred. No. 7.1;
Matches 51; Conservative 28; Mismatches 81; Indels 79; Gaps 9;

QY 227 WEQNTQGSIPYVWLRAPYYWIGRLPSVGHALHEERVERPPMPPT----- 273

Db 342 WERLASSTAPYIPELRGP-----MDTSNFDVDDDTLHNPHTLPPPSHGAFSGHLLPFV 394
QY 274 -FLYTOSWEDPDPMEVMEINPKDVTLTSGGCNALLLVQAGQVSVDCNPAQSALL 332
Db 395 GFTYTSAAALERKLCQLEQ-----KLPAGGSPQLRKEVAALRBOLEQAHSRRLEQEA 448
QY 333 ELKKVAIQ-----LEFEDVMQLFGEVHPRIEELYEKKLAPFLSQTSHNFWSKRLW 384
Db 449 EKOSQALQOELAMLRELEQESKQRL--EGERRETSNWEAQLADILS-----494
QY 385 YFQHGYYQGGMGKLCWV-----LQCLAVVLGLGKTVKRLANAPTMEEQ---RRL 431
Db 495 -----WVNDKVSRCGYLQALAT--KMAEELSRLRVNGTQDHWKARRL 535

RESULT 11

US-10-055-877-10

; Sequence 10, Application US/10055877

; Publication No. US20050288241A1

; GENERAL INFORMATION:

; APPLICANT: DeCristofaro, Marc

; APPLICANT: Padigaru, Muralidhara

; APPLICANT: Miller, Charles

; APPLICANT: Tchernev, Velizar

; APPLICANT: Zhong, Mei

; APPLICANT: Anderson, David

; APPLICANT: Ballinger, Robert

; APPLICANT: Gerlach, Valerie

; APPLICANT: Spytek, Kimberly

; APPLICANT: Ratelli, Luca

; APPLICANT: Kekuda, Ramesh

; APPLICANT: Guo, Xiaojia

; APPLICANT: Zernhusen, Bryan

; APPLICANT: Andrew, David

; APPLICANT: Mezes, Peter

; APPLICANT: Patturajan, Meera

; APPLICANT: Burgess, Catherine

; APPLICANT: Eisen, Andrew

; APPLICANT: Wolenc, Adam

; APPLICANT: Baumgartner, Jason

; APPLICANT: Shimkets, Richard

; APPLICANT: Gusev, Vladimir

; APPLICANT: Vernet, Corine

; APPLICANT: Taupier Jr., Raymond

; APPLICANT: Pena, Carol

; APPLICANT: Shenoy, Suresh

; APPLICANT: Li, Li

; APPLICANT: Casman, Stacie

; TITLE OF INVENTION: Novel Polypeptides and Nucleic Acids Encoded Thereby

; FILE REFERENCE: 21402-251

; CURRENT APPLICATION NUMBER: US/10/055,877

; CURRENT FILING DATE: 2002-01-22

; PRIOR APPLICATION NUMBER: 60/262,892

; PRIOR FILING DATE: 2001-01-19

; PRIOR APPLICATION NUMBER: 60/263,598

; PRIOR FILING DATE: 2001-01-23

; PRIOR APPLICATION NUMBER: 60/263,799

; PRIOR FILING DATE: 2001-01-24

; PRIOR APPLICATION NUMBER: 60/264,117

; PRIOR FILING DATE: 2001-01-25

; PRIOR APPLICATION NUMBER: 60/264,139

; PRIOR FILING DATE: 2001-01-25

; PRIOR APPLICATION NUMBER: 60/264,478

; PRIOR FILING DATE: 2001-01-26

; PRIOR APPLICATION NUMBER: 60/263,351

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: 60/272,870

; PRIOR FILING DATE: 2001-03-02

; PRIOR APPLICATION NUMBER: 60/275,990

; PRIOR FILING DATE: 2001-03-14

; PRIOR APPLICATION NUMBER: 60/275,927

; PRIOR FILING DATE: 2001-03-14

; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 512
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 1247
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-055-877-10

Query Match 2.7%; Score 93.5; DB 6; Length 1247;

Best Local Similarity 21.3%; Pred. No. 7.5; 81; Indels 79; Gaps 9;

Matches 51; Conservative 28; Mismatches 81; Indels 79; Gaps 9;

QY 227 WEONTGSGIPYVWFLRAPPYVVMIGRLPSVGHALHEERVERPPMPPT-----273

Db 342 WERLASSTAPYIPELRGP-----MDTSNFDVDDDTLHNPHTLPPPSHGAFSGHLLPFV 394

QY 274 -FLYTOSWEDPDPMEVMEINPKDVTLTSGGCNALLLVQAGQVSVDCNPAQSALL 332

Db 395 GFTYTSAAALERKLCQLEQ-----KLPAGGSPQLRKEVAALRBOLEQAHSRRLEQEA 448

QY 333 ELKKVAIQ-----LEFEDVMQLFGEVHPRIEELYEKKLAPFLSQTSHNFWSKRLW 384

Db 449 EKOSQALQOELAMLRELEQESKQRL--EGERRETSNWEAQLADILS-----494

QY 385 YFQHGYYQGGMGKLCWV-----LQCLAVVLGLGKTVKRLANAPTMEEQ---RRL 431

Db 495 -----WVNDKVSRCGYLQALAT--KMAEELSRLRVNGTQDHWKARRL 535

RESULT 12

US-11-098-686-11330

; Sequence 11330, Application US/11098686

; Publication No. US20060024696A1

; GENERAL INFORMATION:

; APPLICANT: Kapur, Vivek and Gebhart, Connie J.

; TITLE OF INVENTION: NUCLEIC ACID AND POLYPEPTIDE SEQUENCES

; FILE REFERENCE: 09531-128001

; CURRENT APPLICATION NUMBER: US/11/098,686

; CURRENT FILING DATE: 2005-04-04

; PRIOR APPLICATION NUMBER: PCT/US03/31318

; PRIOR FILING DATE: 2003-10-01

; PRIOR APPLICATION NUMBER: US 60/416,395

; PRIOR FILING DATE: 2002-10-04

; NUMBER OF SEQ ID NOS: 11433

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 11330

; LENGTH: 417

; TYPE: PRT

; ORGANISM: Lawsonia intracellularis

US-11-098-686-11330

Query Match 2.7%; Score 93; DB 7; Length 417;

Best Local Similarity 21.8%; Pred. No. 1.7;

Matches 58; Conservative 43; Mismatches 87; Indels 78; Gaps 16;

QY 56 GPOAAAFARLAERSNLI-----WVDLGGGTGENVDMADYIDLAK-----EKSIVVVDL 105

Db 16 GMTAAAPAAKGGKCLITKNGTTITGSGT---IDVIG-YLPGKPVSNPFAVTFPK 71

QY 106 CHS---LCEVAKKA-----KAKGWKNVQVVEADACQAFPEGT---ATL 144

Db 72 RHPYQLIGELTVRNALKNFSELVGEENVYKTRSQKNYVVT-----AGTIKPTWM 123

QY 145 ITSYSLSMTPTPPHNVNDQACSVLQDGLVGADVFVSGKYDPLRMPMSRRFFWTSIF 204

Db 124 VTESMDMSIIPDIKNI---CIL---GIEGMKDFYP---ELIKGLS-SKKYFANKTF 170

QY 205 DIDNIDIGPER-----AYLEQKLERVW-----EONTGSGIPVWPLRAPPYVWIG 250

Db 171 TQAKLHFPESERNISTNVQAYLETENGLKWLVOGIKQGNQSDLIILPILGTTF-----226

QY 211 IGPERRAYLEQKLERVWEQNTQGSIPVVPWLRAPYVYVWIGRLPSVGHALHEEREVERPPMF 270
Db 202 ---QRRAAAEQEQRLMRQ-----ARMNVDIKLS-----NMSEIRIDSF 240
QY 271 PPTFLYTQSWEDPEPDMVMEINP-----KDTVLTLTSGGCNALNLALVQAGAGQVVSVD--- 323
Db 241 DAVERYL-----LRENE-NPGKGSADSI-----GACQKLKVDLQAAAFNVLPKDLQ 285
QY 324 ---CNPAQSALLE 333
Db 286 LLLSNPKRAVLLQ 298

RESULT 15
US-11-098-686-10262
; Sequence 10262, Application US/11098686
; Publication No. US20060024696A1
; GENERAL INFORMATION:
; APPLICANT: Kapur, Vivek and Gebhart, Connie J.
; TITLE OF INVENTION: NUCLEIC ACID AND POLYPEPTIDE SEQUENCES
; FILE REFERENCE: 09531-128001
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: US/11/098,686
; PRIOR FILING DATE: 2003-10-01
; PRIOR APPLICATION NUMBER: PCT/US03/31318
; PRIOR FILING DATE: 2002-10-04
; NUMBER OF SEQ ID NOS: 11433
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 10262
; LENGTH: 241
; TYPE: PRT
; ORGANISM: Lawsonia intracellularis
US-11-098-686-10262

Query Match 2.68; Score 91.5; DB 7; Length 241;
Best Local Similarity 20.58; Pred. No. 1.1;
Matches 33; Conservative 29; Mismatches 70; Indels 29; Gaps 5;
QY 79 GGTGENVDMMADYIDLA-----KFKSIYV--VDLCHSLCEVAKKAKGKGNQVQVVEAD 131
Db 53 GKTGRILDLAGTLDVAIALHNRVEQLHISAVDLCFSMLYQGMKCLHRANITRWVPTAD 112
QY 132 ACQFAPPEGATLITFSYSLTMTPPFHNVVDQACSYLSQDGLVGVADPFVSGKYDLPRLQ 191
Db 113 AKLLPDPSCIDGVTLAFGIRNIVPRTKAFSEIARVLLPGRMAILEF-GTGKQRI----- 167
QY 192 MPWSRRFFWRSIFD-----IDNIDIGPERRAYLEQKL 223
Db 168 -----WMGLNYLYTKILPFIKLUSSDPSAYLYLKQSI 200

Search completed: March 14, 2006, 02:01:14
Job time : 17 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 14, 2006, 04:22:40 ; Search time 317 Seconds
(without alignments)
3633.628 Million cell updates/sec

Title: US-10-620-914-45

Perfect score: 3463

Sequence: 1 MSGRGRPASVTKNFSLE.....RVNYSSTFNRKGAADN 648

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-O=/abs/ABSSWEB.spool/US10620914/runat_13032006_102046_6830/app.query.fasta_1
-DB=Issued Patents NA -SUFFIX=p2n.rn1 -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCLALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-HOST=abss908 -USBR=US10620914 @CGN 1 1 193 @runat_13032006_102046_6830 -NCPU=6
-ICPU=3 -NO WMAP -NEG SCORES=0 -WAIT -DPPBLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA:
1: /cgn2_6/ptodata/1/ina/1 COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5 COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/1/ina/H COMB.seq.*
6: /cgn2_6/ptodata/1/ina/PCTUS COMB.seq.*
7: /cgn2_6/ptodata/1/ina/PP COMB.seq.*
8: /cgn2_6/ptodata/1/ina/RE COMB.seq.*
9: /cgn2_6/ptodata/1/ina/backfileseq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query No.	Score	Match	Length	ID	Description
1	352	10.2	933	3	US-09-248-796A-7130	Sequence 7130, Ap
2	158	4.6	666	3	US-09-248-796A-12232	Sequence 12232, A
3	138.5	4.0	8832	3	US-09-902-540-984	Sequence 984, App
4	126.5	3.7	990	3	US-09-902-540-9203	Sequence 9203, Ap
5	121	3.5	639	3	US-09-107-532A-3600	Sequence 3600, Ap
6	118.5	3.4	1155	3	US-09-439-313-373	Sequence 373, App
7	118.5	3.4	1155	3	US-09-352-616A-373	Sequence 373, App
8	118.5	3.4	1155	3	US-09-289-198-301	Sequence 301, App
9	118.5	3.4	1155	3	US-09-636-215-373	Sequence 373, App

RESULT 1

US-09-248-796A-7130
; Sequence 7130, Application US/09248796A

; Patent No. 6747137

; GENERAL INFORMATION:

; APPLICANT: Keith Weinstock et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS

; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.132

; CURRENT APPLICATION NUMBER: US/09/248,796A

; CURRENT FILING DATE: 1999-02-12

; PRIOR APPLICATION NUMBER: US 60/074,725

; PRIOR FILING DATE: 1998-02-13

; PRIOR APPLICATION NUMBER: US 60/096,409

; PRIOR FILING DATE: 1998-08-13

; NUMBER OF SEQ ID NOS: 28208

; SEQ ID NO 7130

; LENGTH: 933

; TYPE: DNA

; ORGANISM: Candida albicans

US-09-248-796A-7130

Alignment Scores:

Pred. No.: 6.02e-33 Length: 933
Score: 352.00 Matches: 95
Percent Similarity: 40.5% Conservative: 35
Best Local Similarity: 29.6% Mismatches: 97
Query Match: 10.2% Indels: 94
DB: 3 Gaps: 8

US-10-620-914-45 (1-648) x US-09-248-796A-7130 (1-933)

QY

144 Leu1ThrPheSerTyrSerLeuThrMetIleProPheHisnValIleAppGln 163

Query Match: 4.0% Indels: 107
DB: 3 Gaps: 17

US-10-620-914-45 (1-648) x US-09-902-540-984 (1-8832)

QY 252 LeuProSerValGlyHisAlaLeuHisGluArgValGluArgProProMetPhePro 271
DB 2341 GTTCCGAGGTAGAACCATGCAATGGAACCTGGCTGGTTTCGAGCGCGCCCTT-----2288

QY 272 ProThrPheLeuTyrThrGlnSerTrpGluAspProGluProAspMetGluValMetGlu 291
DB 2287 ---CGCTTGAAGTTGCGGCTGTCGCGGAGGATGCGCGCTGGAGTTGGCGCTCGTCGAG 2231

QY 292 IleAsnProLysAspThrValLeuThrLeuThrSerGlyGlyCysAsnAlaLeuAsnLeu 311
DB 2230 CGCACGCGCGCGCGCGCTGCTCACGCTGGCTCGCGAGGTTGCACACTCTCACCTG 2171

QY 312 LeuValGln---GlyAlaGlyGlnValValSerValAspCysAsnProAlaGlnSerAla 330
DB 2170 GCGCGGCGCATCAGCGCTCGAGCTGGTGGCTTCGATTTCACCGCGCTCAGTCGCG 2111

QY 331 LeuLeuGluLeuLysValAlaAlaGlnGlnLeu-----Glu 343
DB 2110 CATGTCGCGGAGAGCGGAGGCTCGCTCGCTCGCGCTCGCTGCGCGCTACAGCGTGGAT 2051

QY 344 PheGluAspValTrpGlnLeu-----PheGlyGluGlyValHisProArg 358
DB 2050 GCGGAGGACGCGCGCGCTGGAATCAGCGCGCGCAATTC---GAGGCGCTCTTCCGACG 1994

QY 359 IleGluLeuLeuTyrGluLysLysLeuAlaProPheLeuSerGlnThrSerHis-AsnPh 378
DB 1993 CTCGCGCGCTTCATCGAGGAGTTCGTCGCGCC-----GCGCATGAAC-- 1951

QY 378 eTrpSerLysArg-LeuTrpTyrPheGlnHisGlyLeuTyrTyrGlnGlyMetGlyL 398
DB 1950 ----TCGCGCGCTTCTCGCGCGCGCACCGGCTCACACGCGGAGGC----- 1902

QY 398 yLeuCysTrpValLeuGlnCysLeuAlaValLeuGlyLeuGlyLysThrValLysA 418
DB 1901 --CTGTGC----- 1896

QY 418 rGleuAlaAsnAlaProThrMetGluGluGlnArgLeuTrpAspSerAsnMetLeuI 438
DB 1895 -----CCGCTGG 1889

QY 438 leHisPheValLysAsnGlyProLysProLeuValTrpLeuPheValLysPheValSerL 458
DB 1888 TTCGCTTGGCCT-ACTGGCCGCTGGCTTCGAGCTGGCTCGCGGCGCG-CTGCTGA 1831

QY 458 euValLeuPheAsnLysAlaValLeuTrpPheGly-----GlyGlyValProGlyLysG 476
DB 1830 ACACCATGTTTGTTCGCGCGCGCACGATCGGAGCGCGCTCGTATCTCTGGCTATT 1771

QY 476 lNTrAlaLeuIleLysAlaAspGlyIleProIleGluAsnTyrIleAlaArgThrMetA 496
DB 1770 TCAGGCGGCTTCGAACGC---GGGCTCCAGCGCGAGGAT----- 1733

QY 496 spGlyValAlaGluAsnSerHisValArgLysGlnAsnTyrPheTyrTyrAsnCysLeu 516
DB 1732 -----GCGCTCGAATCCCTTCCTCCAGCAGCTGCTGC 1699

QY 516 hrGlyLysPheLeuArgAspAsnCysProThrTyrLeuArg---GluAlaAlaPheAla- 534
DB 1698 TCGGAGGTACCTCGCTGAGGATGCCCCGAGTATCTCCGGGCTGAAGGCGCGCTGCGCG 1639

QY 535 ---ThrLeuLysSerGlyValValAspAsnLeuThrValSerThrAsnPheMetGluG 554
DB 1638 TGACGCTCGTCCAGGCGCTGTTCCCGGATGTGCGCGCTCGGACCGCTTC----- 1589

QY 554 luLeuLysAlaArgThrTyrThrLysValIleLeuMetAspHisVal---AspTrpLeuA 573
DB 1588 -----GAGCTCATCTCGCTCTCCAAACATCTTCGACTGTGCTCGG 1552

QY 573 spMetProValAlaAsnGluLeuAlaGluCysLeuAlaLysGlnValAlaProGlyGlyI 593
DB 1551 AAGACGCGCTGTGGCGGAGTGGCGGGTGTCTGCACGCGCGCGCGCTTTCGAGCGCGCT 1492

QY 593 leValIleTrpArg-Ser-----AlaSerLeuSerProPro 604
DB 1491 CCGTCTCATCGCGAGCTCAACACGCGCGGACCTTCGCGGCTTCTTCAGCGCGCT 1432

QY 605 TyrAlaGluLeuLeuGlnLysAla 612
DB 1431 TCGAGTTCGATGACGCGCTCGCGG 1408

RESULT 4
US-09-902-540-9203
; Sequence 9203, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 9203
; LENGTH: 990
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-9203

Alignment Scores:
Pred. No.: 4,18e-05 Length: 990
Score: 126.50 Matches: 94
Percent Similarity: 37.6% Conservative: 48
Best Local Similarity: 24.9% Mismatches: 131
Query Match: 3.7% Indels: 107
DB: 3 Gaps: 17

US-10-620-914-45 (1-648) x US-09-902-540-9203 (1-990)

QY 264 ValGluArgProProMetPheProThrPheLeuTyrThrGlnSerTrpGluAspPro 283
DB 16 GTTTCGACGCGCCCTT-----CGCTTGAAGTTCGCGCTTCGCGGAGGATGCC 66

QY 284 GluProAspMetGluValMetGluLleAsnProLysAspThrValLeuThrLeuThrSer 303
DB 67 GCGCTGGAGTTGGCTGCTCGACGCGCACGCGGCGCGCTGCTCAGGTGGCGCTCG 126

QY 304 GlyGlyCysAsnAlaLeuAsnLeuValGln---GlyAlaGlyGlnValValSerVal 322
DB 127 GGAGGTTCGACACTCTCTACCTTGGCGCGCGCCATCCAGCTCGAGCTGGTGGGCTTC 186

QY 323 AspCysAsnProAlaGlnSerAlaLeuLeuGluLeuLysValAlaIleGlnGlnLeu 342
DB 187 GATTTCACCCCGCTCAGCTCGCGCATGTCGCGGAGAGCGGAGGCGGCTCGCTCGCT 246

QY 343 -----GluPheGluAspValTrpGlnLeu----- 350
DB 247 CCGCTGGCGCTACAGCTGATGATCGGAGGACGCGCGGCTGATCAGCGCGCGAA 306

QY 351 PheGlyGluGlyValHisProArgIleGluLeuLeuTyrGluLysLysLeuAlaProPhe 370
DB 307 TTC---GAGGCGCTCTTCGCGACGCTCCCGCGCTTCATCGAGGATTCGTCGCGCC--- 360

QY 371 LeuSerGlnThrSerHis-AsnPheTrpSerLysArg-LeuTrpTyrPheGlnHisGlyL 390
DB 361 -----GCGCATGAAC-----TCGCGCGCTTCTTCGCGCGCGCGCACCGCC 402

QY 390 euTyrTyrGlnGlyMetGlyLysLeuCysTrpValLeuGlnCysLeuAlaValI 410

```
Db 403 TCACACGCGCGGAGGC-----CTGTGC----- 425
Qy 410 euGlyLeuGlyLeuThrValysArgLeuAlaAsnAlaProThrMetGluGluGlnArgA 430
Db 425 ----- 425
Qy 430 rgLeuTrpAspSerAsnMetLeuIleHisPheValysAsnGlyProLysProLeuValt 450
Db 426 -----CCGCTGGTTCGCTTCGCCT-CTGGCCCGTGGCCCTTCGAGCT 467
Qy 450 rpLeuPheValysPheValSerLeuValLeuPheAsnLysAlaValLeuTrpPheGly- 469
Db 468 GGCCCTGGCGCGCG-CTGCTGAACACCATGTTTGGTCCGCGCGGACCCAGCATGCCG 526
Qy 470 -----GlyGlyValProGlyGlnTrpAlaLeuIleLysAlaAspGlyIleProileg 488
Db 527 AGCCCGGCTCGTATCTTCGCGGTATTTCCAGCGGTCTTCGAAACG----GGCTCCAGCGCG 583
Qy 488 luAsnTrpIleAlaArgThrMetAspGlyValAlaGluAsnSerHisValArgLysGlnA 508
Db 584 AGGAT-----GCGCTCGAA 598
Qy 508 snTrpPheTyrTrpAsnCysLeuThrGlyLysPheLeuArgAspAsnCysProThrTyrL 528
Db 599 ATCCCTTCCTCCAGCACGCTGCTCGCGCAGGTACCTGCGTGAAGATGCCCGGAGTATC 658
Qy 528 euArg-----GluAlaAlaPheAla---ThrLeuLysSerGlyValValAspAsnLeuThrV 546
Db 659 TCCGGGTGAAGGCCCTCGCGCTGACGCTGCCAGCGCTGTCGCGGATGTGCCG 718
Qy 546 alSerThrAsnPhePheMetGluGluLeuLysAlaArgThrTyrThrLysValIleLeuM 566
Db 719 GCCTGGACCGCTTC-----GACGTCACTCCGC 745
Qy 566 etAspHisVal---AspTrpLeuAspMetProValAlaAsnGluLeuAlaGluCysLeuA 585
Db 746 TCTCCACATCTTCGACTGTCGGAAGACGCGCTGTCGCGGAGTGGCGGGGTGCTCG 805
Qy 585 lalysGlnValAlaProGlyGlyIleValIleTrpArg-Ser----- 598
Db 806 CAGGTGAGCGCGCGCGTGGCGCTCTCTCATCCCGCCAGCTCAACACCGCGCGGACC 865
Qy 599 -----AlaSerLeuSerProTrpTyrAlaGluLeuIleGlnLysAla 612
Db 866 TTCGCGCTTCCTCCAGCCCGCTTCGAGTTCGATGACGCGCTCGCG 913

RESULT 5
US-10-532A-3600
; Sequence 3600, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
```

```
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 3600:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 639 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...639
; SEQUENCE DESCRIPTION: SEQ ID NO: 3600:
US-09-107-532A-3600

Alignment Scores:
Pred. No.: 9,28e-05 Length: 639
Score: 121.00 Matches: 33
Percent Similarity: 51.9% Conservative: 23
Best Local Similarity: 30.6% Mismatches: 44
Query Match: 3.5% Indels: 8
DB: 3 Gaps: 3

US-10-620-914-45 (1-648) x US-09-107-532A-3600 (1-639)
Qy 73 IleTrpValAspLeuGlyGlyThrGlyGluAsnValAspMetMetAlaAspTyrIle 92
Db 157 GTCTGCTGTGATTACGAGGCGGAACTGGGCTTTAGTTTACGCTTGGGAA----- 210
Qy 93 AspLeuAlalysPheLysSerIleTyrValValAspLeuCysHisSerLeuCysGluVal 112
Db 211 -----AGATTCAAAGAACTGATCATTCGCGATGCTTCAGAAACTATGCTGAAATG 261
Qy 113 AlalysLysLysAlalysAlalysGlyTyrLysAsnValGlnValGluAlaAspAla 132
Db 262 GCAGAGAAAGATTCAGCAGCAGATCTGAAAAATGTCGAGACGATCCACGCTGATGCT 321
Qy 133 ---CysGlnPheAlaProGluGlyThrAlaThrLeuIleThrPheSerTyrSerLeu 151
Db 322 TCTGTGCAATTT-----CCTGCTGTCCAAGCTAATCTCATCTCTTCTCATCTGTACTT 375
Qy 152 ThrMetIleProPheHisAsnValIleAspGlnAlaCysSerTyrLeuSerGlnAsp 171
Db 376 TTACACATTCCTGATACGAGAGATATTCTTACAAACTTTTATGAAATTTTAGTCTCTGGC 435
Qy 172 GlyLeuValGlyValAlaAspPhe 179
Db 436 GGTCATTTGATCATCTGATGACTTT 459

RESULT 6
US-09-439-313-373
; Sequence 373, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang Yuqi
; APPLICANT: Reed, Steven G.
; APPLICANT: Kalos, Michael
```

187	Qy	LeuProLeuArgGlnMetProTpr-----SerArgArgPhePheTrpArgSerIlePhe	204
204	Db	CTTCCCCTGTCGAGGGGAGTGCACAGACACGTCGGCGCTTCTCGAGA-----	254
205	Qy	AspIleAspAsnIleAspIleGlyProGluArgArgAlaTyrLeuGluGlnLysLeuGlu	224
255	Db	-----CCACGACGACTCTGCTAT-----GAA	275
225	Qy	ArgValTrpGluGlnAsnThrGlnGlySerIleProTyrValProTrpLeuArgAlaPro	244
276	Db	GACACTCAGAAACAGATGGCGAGTGGTCTGCCACTGCTCCCTCGTCGAGGGGAG	335
245	Qy	TyrTyrValTrpIleGlyArgLeuProSerValGly---HisAlaLeuHisGluGluArg	263
336	Db	CGGCAAGACGAGTGGCGCTTGGGAGACTACGATGACAGTCGCTTCATGGAGCCGAG	395
264	Qy	ValGluArgProProMetPheProProThrPheLeuTyrThrGlnSerTrp-GluAsp--	282
396	Db	GTACCAGTCCG-----TGCAGAGAATCT	419
283	Qy	-----ProGluProAspMetGluValMe	290
420	Db	GGCAAGACTCCACAGAGCTCCCTGGTGGGGTAAAGTCCCCAGAAAGGATCTCATCGTCTAT	479
290	Qy	t-----GluIleAsnProLysAsp-----ThrValLeuThrLeuTh	302
480	Db	GCTCAGGGACACTGGACGTGAACAGAGGACAAAGCAAAAGAGGACTGCTCTACATCTGCG	539
302	Qy	rSer-----GlyGlyCysAsnAlaLeuAsnLeuValGlnGlyAlaGlyGlnValVa	320
540	Db	CTCTGCCAATGGGAATTCAAGAAGTAGTAAAACTCCTGCTGGACAGACAGATGTCAACTTAA	599
320	Qy	lSerValAspCysAsnProAlaGlnSerAlaLeuLeuGluLeuLysLysValAlaIleGl	340
600	Db	TGTCCTTGAC--AACAAAAAGGACAGCTCTGATAAG-----GCCGTACA	644
340	Qy	nGlnLeuGluPheGluAspValTrpGlnLeuPheGlyGluGlyValHisProArGileGl	360

```

Db      645 ATCCAGGAAGATGAATGTCGTTAAATGTTGTCGGAACATGGCACTGATCAAAATATTC 704
QY      360 uGluLeuTyr-GluLysLysLeuAlaProPheLeuSerGlnThrSerHisAsnPheTrpS 380
Db      705 AGATGAGTATGGAATACCACTCTGCACACTACGCTATCTATAATGAAGATAAAATTAATG 764
QY      380 erLysArgLeuTrpTyr 385
Db      765 CAAGCACTCTCTTAT 781

RESULT 8
US-09-289-198-301
; Sequence 301, Application US/09289198
; Patent No. 6586570
; GENERAL INFORMATION:
; APPLICANT: Fridakis, Tony N.
; APPLICANT: Smith, John M.
; APPLICANT: Reed, Steven G.
; APPLICANT: Misher, Lynda
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TREATMENT OF BREAST CANCER
; FILE REFERENCE: 210121.419CS
; CURRENT APPLICATION NUMBER: US/09/289,198
; CURRENT FILING DATE: 1999-04-09
; EARLIER APPLICATION NUMBER: US 09/062,451
; EARLIER FILING DATE: 1998-04-17
; EARLIER APPLICATION NUMBER: US 08/991,789
; EARLIER FILING DATE: 1997-12-11
; EARLIER APPLICATION NUMBER: US 08/838,762
; EARLIER FILING DATE: 1997-04-09
; EARLIER APPLICATION NUMBER: PCT/US97/00485
; EARLIER FILING DATE: 1997-01-10
; EARLIER APPLICATION NUMBER: US 08/700,014
; EARLIER FILING DATE: 1996-08-20
; EARLIER APPLICATION NUMBER: US 08/585,392
; EARLIER FILING DATE: 1996-01-01
; NUMBER OF SEQ ID NOS: 312
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 301
; LENGTH: 1155
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-289-198-301

Alignment Scores:
Pred. No.: 0.000529 Length: 1155
Score: 118.50 Matches: 66
Percent Similarity: 37.4% Conservative: 19
Best Local Similarity: 29.1% Mismatches: 81
Query Match: 3.4% Indels: 61
DB: 3 Gaps: 11

US-10-620-914-45 (1-648) x US-09-289-198-301 (1-1155)
QY      187 LeuProLeuArgGlnMetProTrp-----SerArgArgPhePheTrpArgSerIlePhe 204
Db      204 CTTCCTCTCTCAGGGGAGTGGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 254
QY      205 AspIleAsnIleAspIleGlyProGluArgAlaTyrLeuGluGlnLysLeuGlu 224
Db      255 -----CCACGACACTCTGCTAT-----GAA 275
QY      225 ArgValTrpGluGlnAsnThrGlnGlySerIleProTyrValProTyrLeuArgAlaPro 244
Db      276 GACACTCAGGAACAGATGGCAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 335
QY      245 TyrTyrValTrpIleGlyArgLeuProSerValGly---HisAlaLeuHisGluGluArg 263
Db      336 CGGCAGAGCAAGGTGGCGCTTGGGAGACTACGATCAGATGAGTGGCTTCATGGAGCCAG 395
QY      264 ValGluArgProProMetPheProThrPheLeuTyrThrGlnSerTrp-GluAsp-- 282
Db      396 GTACCACGTCG-----TGCAGAGATCT 419

```

```

QY      283 -----ProGluProAspMetGluValme 290
Db      420 GGCAAGCTCCACAGAGCTGCCTGGTGGGTAAAGTCCCAAGAAAGATCTCATCGTCAT 479
QY      290 t-----GluIleAsnProLysAsp-----ThrValLeuThrLeuTh 302
Db      480 GCTCAGGGACACTGACGTGAACAAGAGGACAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 539
QY      302 rSer-----GlyGlyCysAsnAlaLeuAsnLeuValGlnGlyGlnValVa 320
Db      540 CTCTGCCAATGGGAATTCAGAACTAGTAAACTCTGCTGGACAGAGGATGTCACATTAA 599
QY      320 lSerValAspCysAsnProAlaGlnSerAlaLeuLeuGluLeuLysLysValAlaIleGl 340
Db      600 TGTCTTGTGAC--AACAAAAAGAGGACAGCTCTGATAAAG-----GCCGTACA 644
QY      340 nGlnLeuGluPheGluAspValTrpGlnLeuPheGlyGlyGluGlyValHisProArgIleGl 360
Db      645 ATGCCAGGAAGATGAATGTCGTTAATGTTGCTGGAACTGTCGCACTGATCCTCAATATTC 704
QY      360 uGluLeuTyr-GluLysLysLeuAlaProPheLeuSerGlnThrSerHisAsnPheTrpS 380
Db      705 AGATGAGTATGGAATACCACTCTGCACACTACGCTATCTATAATGAAGATAAAATTAATG 764
QY      380 erLysArgLeuTrpTyr 385
Db      765 CAAGCACTCTCTTAT 781

```

RESULT 9

```

US-09-636-215-373
; Sequence 373, Application US/09636215
; Patent No. 6620922
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqi
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TREATMENT OF PROSTATE CANCER
; FILE REFERENCE: 210121.42717C17
; CURRENT APPLICATION NUMBER: US/09/636,215
; CURRENT FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 852
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 373
; LENGTH: 1155
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-636-215-373

```

```

Alignment Scores:
Pred. No.: 0.000529 Length: 1155
Score: 118.50 Matches: 66
Percent Similarity: 37.4% Conservative: 19
Best Local Similarity: 29.1% Mismatches: 81
Query Match: 3.4% Indels: 61
DB: 3 Gaps: 11

```

US-10-620-914-45 (1-648) x US-09-636-215-373 (1-1155)

GENERAL INFORMATION:
; APPLICANT: Fridakis, Tony N.
; APPLICANT: Smith, John M.
; APPLICANT: Reed, Steven G.
; APPLICANT: Misher, Lynda
; APPLICANT: Retter, Marc W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TREATMENT AND DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.419C6
; CURRENT APPLICATION NUMBER: US/09/429,755A
; CURRENT FILING DATE: 1999-10-28
; NUMBER OF SEQ ID NOS: 315
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 301
; LENGTH: 1155
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-429-755-301

Alignment Scores:
Pred. No.: 0.000529 Length: 1155
Score: 118.50 Matches: 66
Percent Similarity: 37.4% Conservative: 19
Best Local Similarity: 29.1% Mismatches: 81
Query Match: 3.4% Indels: 61
DB: 3 Gaps: 11

US-10-620-914-45 (1-648) x US-09-429-755-301 (1-1155)

```
QY 187 LeuProLeuArgGlnMetProTTP-----SerArgArgPhePheTTPArgSerIlePhe 204
Db 204 CTTCCTCCCTGCTGCAGGGGGAGTGGCAAGACAGAGCTGGGGCGCTTCTGAGA----- 254

QY 205 AspIleAspAsnIleAspIleGlyProGluArgAlaTyrLeuGluGlnLysLeuGlu 224
Db 255 -----CCACGACGACTCTGCTAT-----GAA 275

QY 225 ArgValTrpGluGlnAsnThrGlnGlySerIleProTyrValProTyrLeuArgAlaPro 244
Db 276 GACACTCAGGAACAAGATGGCAAGAGTGGTGCCTGCCACTGTCCCTGCTGCAGGGGGAG 335

QY 245 TyrTyrValTrpIleGlyArgLeuProSerValGly---HisAlaLeuHisGluGluArg 263
Db 336 CGGCAAGACGACAGTGGGGCGCTTGGGAGACTACGATGACAGTCCCTTCATGGAGCCAG 395

QY 264 ValGluArgProProMetPheProProThrPheLeuTyrThrGlnSerTrp-GluAsp-- 282
Db 396 GTACCACGTCCG-----TGAGAGAAGATCT 419

QY 283 -----ProGluProAspMetGluValMe 290
Db 420 GGACACAGCTCCACAGAGCTGCCTGGTGGGGTAAAGTCCCCAGAAAGGATCTCTCGTCAT 479

QY 290 t-----GluIleAsnProLysAsp-----ThrValLeuThrLeuTh 302
Db 480 GCTCAGGACACTGACGTGAACAAGAGGACAGCAACAAAGAGGACTGCTCTACATCTGGC 539

QY 302 rSer-----GlyGlyCysAsnAlaLeuAsnLeuLeuValGlnGlyAlaGlyGlnValVa 320
Db 540 CTCTGCCAATGGGAATTGAGAAGTAGTAAAACTCCTGCTGGACAGACGATGTCAACTTAA 599

QY 320 lSerValAspCysAsnProAlaGlnSerAlaLeuLeuGluLeuLysValAlaIleGl 340
Db 600 TGCTCTTGAC---AACAAGAGGACGACTCTGATAAAG-----GCCGTACA 644

QY 340 nGlnLeuGluPheGluAspValTrpGlnLeuPheGlyGluGlyValHisProArgIleGl 360
Db 645 ATGCAGGACAGATGAATGTGGTTAATGTGCTGGAACATGGCACTGATCCAAATATTC 704

QY 360 uGluLeuTyr-GluLysLysLeuAlaProPheLeuSerGlnThrSerHisAsnPheTrps 380
Db 705 AGATGAGTATGGAATACCACTCTGCACTACGCTATCTATAATGAAGATAAATTAATGGC 764
```

```
QY 380 erLysArgLeuTrpTyr 385
Db 765 CAAGCACTGCTCTTAT 781

RESULT 12
US-09-679-426-373
; Sequence 373, Application US/09679426
; Patent No. 679515
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqui
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TREATMENT OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C20
; CURRENT APPLICATION NUMBER: US/09/679,426
; CURRENT FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 895
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 373
; LENGTH: 1155
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-679-426-373
```

Alignment Scores:
Pred. No.: 0.000529 Length: 1155
Score: 118.50 Matches: 66
Percent Similarity: 37.4% Conservative: 19
Best Local Similarity: 29.1% Mismatches: 81
Query Match: 3.4% Indels: 61
DB: 3 Gaps: 11

US-10-620-914-45 (1-648) x US-09-679-426-373 (1-1155)

```
QY 187 LeuProLeuArgGlnMetProTTP-----SerArgArgPhePheTTPArgSerIlePhe 204
Db 204 CTTCCTCCCTGCTGCAGGGGGAGTGGCAAGACAGAGCTGGGGCGCTTCTGAGA----- 254

QY 205 AspIleAspAsnIleAspIleGlyProGluArgAlaTyrLeuGluGlnLysLeuGlu 224
Db 255 -----CCACGACGACTCTGCTAT-----GAA 275

QY 225 ArgValTrpGluGlnAsnThrGlnGlySerIleProTyrValProTyrLeuArgAlaPro 244
Db 276 GACACTCAGGAACAAGATGGCAAGAGTGGTGCCTGCCACTGTCCCTGCTGCAGGGGGAG 335

QY 245 TyrTyrValTrpIleGlyArgLeuProSerValGly---HisAlaLeuHisGluGluArg 263
Db 336 CGGCAAGACGACAGTGGGGCGCTTGGGAGACTACGATGACAGTCCCTTCATGGAGCCAG 395

QY 264 ValGluArgProProMetPheProProThrPheLeuTyrThrGlnSerTrp-GluAsp-- 282
Db 396 GTACCACGTCCG-----TGAGAGAAGATCT 419

QY 283 -----ProGluProAspMetGluValMe 290
Db 420 GGACACAGCTCCACAGAGCTGCCTGGTGGGGTAAAGTCCCCAGAAAGGATCTCTCGTCAT 479
```

Qy 290 t-----GluLeuAsnProLysAsp-----ThrValLeuThrLeuTh 302
Db 480 GCTCAGGACACTGACGTGAACAAGAGGCAAGCAAAAGAGGACTGCTTACATCTGGC 539
Qy 302 rSer-----GlyGlyCysAsnAlaLeuAsnLeuValGlnGlyAlaGlyGlnValva 320
Db 540 CTCTGCCAATGGGAATTCAGAGTAGTAAACTCTCTGTCGACAGACGATGTCAACTTAA 599
Qy 320 lSerValAspCysAsnProAlaGlnSerAlaLeuLeuGluLeuLysValAlaIleG1 340
Db 600 TGTCTTTGAC---AACAAAAGAGGACAGCTCTGATAAG-----GCCGTACA 644
Qy 340 nGlnLeuGluPheGluAspValTrpGlnLeuPheGlyGluGlyValHisProArgIleG1 360
Db 645 ATGCAGGAAGATGAATGTGGTTAATGTGTGTAACATGGCAGTCTGATCCAAATATTC 704
Qy 360 uGluLeuTyr-GluLysLysLeuAlaProPheLeuSerGlnThrSerHisAsnPheTrpS 380
Db 705 AGATGAGTATGGAATACCACTCTGCACCTAGCTATCTATAATGAAGATAAATAATGGC 764
Qy 380 erLysArgLeuTrpTyr 385
Db 765 CAAAGCACTGCTCTTAT 781

RESULT 13

US-09-759-143-373
; Sequence 373, Application US/09759143
; Patent No. 6800746

GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqui
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND

; FILE REFERENCE: 210121.427C23

; CURRENT APPLICATION NUMBER: US/09/759,143

; CURRENT FILING DATE: 2001-01-12

; NUMBER OF SEQ ID NOS: 934

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 373

; LENGTH: 1155

; TYPE: DNA

; ORGANISM: Homo sapien

US-09-759-143-373

Alignment Scores:

Pred. No.:	0.000529	Length:	1155
Score:	118.50	Matches:	66
Percent Similarity:	37.4%	Conservative:	19
Best Local Similarity:	29.1%	Mismatches:	81
Query Match:	3.4%	Indels:	61
DB:	3	Gaps:	11

US-10-620-914-45 (1-648) x US-09-759-143-373 (1-1155)

Qy 187 LeuProLeuArgGlnMetProTrp-----SerArgArgPheTrpArgSerIlePhe 204
Db 204 CTTCCCTGTCGAGGGGAGTGGCAAGACGTGGGGCTTCTGGAGA----- 254

Qy 205 AspIleAspAsnIleAspIleGlyProGluArgAlaTyrLeuGluGlnLysLeuGlu 224
Db 255 -----CCACGACGACTCTGCTAT-----GAA 275
Qy 225 ArgValTrpGluGlnAsnThrGlnGlySerIleProTyrValProTrpLeuArgAlaPro 244
Db 276 GACACTCAGGAAACAAGATGGGCAAGTGGTCTGCCACTGCTTCCCTCGCGAGGGGAG 335
Qy 245 TyrTyrValTrpIleGlyArgLeuProSerValGly---HisAlaLeuHisGluGluArg 263
Db 336 CGCAAGACCAAGCTGGCGCTTGGGGAGACTACGATGACAGTGCCTTCTATGAGGCCAG 395
Qy 264 ValGluArgProProMetPheProProThrPheLeuTyrThrGlnSerTrp-GluAsp-- 282
Db 396 GTACCAGCTCCG-----TGGAGAAGATCT 419
Qy 283 -----ProGluProAspMetGluValme 290
Db 420 GGACAAGCTCCACAGAGCTGCCTGGTGGGTAAAGTCCCAAGAAAGGATCTCATCGCAT 479
Qy 290 t-----GluIleAsnProLysAsp-----ThrValLeuThrLeuTh 302
Db 480 GCTCAGGACACTGACGTGAACAAGAGGCAAGCAAAAGAGGACTGCTTACATCTGGC 539
Qy 302 rSer-----GlyGlyCysAsnAlaLeuAsnLeuValGlnGlyAlaGlyGlnValva 320
Db 540 CTCTGCCAATGGGAATTCAGAGTAGTAAACTCTCTGTCGACAGACGATGTCAACTTAA 599
Qy 320 lSerValAspCysAsnProAlaGlnSerAlaLeuLeuGluLeuLysValAlaIleG1 340
Db 600 TGTCTTTGAC---AACAAAAGAGGACAGCTCTGATAAG-----GCCGTACA 644
Qy 340 nGlnLeuGluPheGluAspValTrpGlnLeuPheGlyGluGlyValHisProArgIleG1 360
Db 645 ATGCAGGAAGATGAATGTGGTTAATGTGTGTAACATGGCAGTCTGATCCAAATATTC 704
Qy 360 uGluLeuTyr-GluLysLysLeuAlaProPheLeuSerGlnThrSerHisAsnPheTrpS 380
Db 705 AGATGAGTATGGAATACCACTCTGCACCTAGCTATCTATAATGAAGATAAATAATGGC 764
Qy 380 erLysArgLeuTrpTyr 385
Db 765 CAAAGCACTGCTCTTAT 781

RESULT 14

US-09-651-236-373

; Sequence 373, Application US/09651236

; Patent No. 6818751

; GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqui
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND

; FILE REFERENCE: 210121.42718C18

; CURRENT APPLICATION NUMBER: US/09/651,236

; CURRENT FILING DATE: 2000-08-29

; NUMBER OF SEQ ID NOS: 865

; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 373
; LENGTH: 1155
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-651-236-373

Alignment Scores:
Pred. No.: 0.000529 Length: 1155
Score: 118.50 Matches: 66
Percent Similarity: 37.4% Conservative: 19
Best Local Similarity: 29.1% Mismatches: 81
Query Match: 3.4% Indels: 61
DB: 3 Gaps: 11

US-10-620-914-45 (1-648) x US-09-651-236-373 (1-1155)

```
Qy 187 LeuProLeuArgGlnMetProTrp-----SerArgArgPhePheTrpArgSerIlePhe 204
Db 204 CTTCCCTGCTGCAGGGGGAGTGGCAAGAGCAACGTGGGGCTCTCGAGA-----254
Qy 205 AspIleAspAsnIleAspIleGlyProGluArgAlaTyrlLeuGluGlnLysLeuGlu 224
Db 255 -----CCACGACGACTCTGCTAT-----GAA 275
Qy 225 ArgValTrpGluGlnAsnThrGlnGlySerIleProTyrlValProTrpLeuArgAlaPro 244
Db 276 GACACTCAGGACACAGATGGCAAGTGGTGTGCTGCCACTGCTTCCCTGCTGCAGGGGGAG 335
Qy 245 TyrTyrlValTrpIleGlyArgLeuProSerValGly---HisAlaLeuHisGluGluArg 263
Db 336 CGGCAAGAGCAAGTGGGGCTTGGGGAGACTACGATGACAGTGCCTTCTATGAGCCAG 395
Qy 264 ValGluArgProProMetPheProThrPheLeuTyrlThrGlnSerTrp-GluAsp-- 282
Db 396 GTACCACGCTCG-----TGAGAAGATCT 419
Qy 283 -----ProGluProAspMetGluValMe 290
Db 420 GGACNAGCTCCACAGAGCTGCTGGTGGGTAAAGTCCCGAAGAGATCTCATCGTCAT 479
Qy 290 t-----GluIleAsnProLysAsp-----ThrValLeuThrLeuTh 302
Db 480 GCTCAGGACACTGACCTGACGTAACAAGAGGACAAAGAGGACTGCTCTACATCTGGC 539
Qy 302 rSer-----GlyGlyCysAsnAlaLeuAsnLeuValGlnGlyAlaGlyGlnValVa 320
Db 540 CTCTGCCAATGGGAATTCAGAAAGTAGTAAACCTCTCTGGACAGACGATGTCAACTTAA 599
Qy 320 lSerValAspCysAsnProAlaGlnSerAlaLeuLeuGluLeuLysLysValAlaIleGl 340
Db 600 TGTCCTTGAC---AACAAGAGGACAGCTCTGATAAG-----GCCGTACA 644
Qy 340 nGlnLeuGluPheGluAspValTrpGlnLeuPheGlyGluGlyValHisProArgIleGl 360
Db 645 ATGCCAGGAGATGAATGCTGTTAATGTTGCTGGACATGGCACTGATCAATAATTC 704
Qy 360 uGluLeuTyrl-GluLysLysLeuAlaProPheLeuSerGlnThrSerHisAsnPheTrpS 380
Db 705 AGATGAGTATGGAATACCACTCTGCCTACTAGCTATCTATAATGAAGATAAATTAATGGC 764
Qy 380 erLysArgLeuTrpTyrl 385
Db 765 CAAGCAGCTGCTCTTAT 781
```

RESULT 15

US-09-699-295-301
; Sequence 301, Application US/09699295
; Patent No. 5828431
; GENERAL INFORMATION:
; APPLICANT: Frudakis, Tony N.
; APPLICANT: Reed, Steven G.
; APPLICANT: Smith, John M.

; APPLICANT: Misher, Linda E.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Retter, Marc W.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; THERAPY AND DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.419C10
; CURRENT APPLICATION NUMBER: US/09/699,295
; CURRENT FILING DATE: 2000-10-26
; NUMBER OF SEQ ID NOS: 326
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 301
; LENGTH: 1155
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-699-295-301

Alignment Scores:
Pred. No.: 0.000529 Length: 1155
Score: 118.50 Matches: 66
Percent Similarity: 37.4% Conservative: 19
Best Local Similarity: 29.1% Mismatches: 81
Query Match: 3.4% Indels: 61
DB: 3 Gaps: 11

US-10-620-914-45 (1-648) x US-09-699-295-301 (1-1155)

```
Qy 187 LeuProLeuArgGlnMetProTrp-----SerArgArgPhePheTrpArgSerIlePhe 204
Db 204 CTTCCCTGCTGCAGGGGGAGTGGCAAGAGCAACGTGGGGCTCTCGAGA-----254
Qy 205 AspIleAspAsnIleAspIleGlyProGluArgAlaTyrlLeuGluGlnLysLeuGlu 224
Db 255 -----CCACGACGACTCTGCTAT-----GAA 275
Qy 225 ArgValTrpGluGlnAsnThrGlnGlySerIleProTyrlValProTrpLeuArgAlaPro 244
Db 276 GACACTCAGGACACAGATGGCAAGTGGTGTGCTGCCACTGCTTCCCTGCTGCAGGGGGAG 335
Qy 245 TyrTyrlValTrpIleGlyArgLeuProSerValGly---HisAlaLeuHisGluGluArg 263
Db 336 CGGCAAGAGCAAGTGGGGCTTGGGGAGACTACGATGACAGTGCCTTCTATGAGCCAG 395
Qy 264 ValGluArgProProMetPheProThrPheLeuTyrlThrGlnSerTrp-GluAsp-- 282
Db 396 GTACCACGCTCG-----TGAGAAGATCT 419
Qy 283 -----ProGluProAspMetGluValMe 290
Db 420 GGACNAGCTCCACAGAGCTGCTGGTGGGTAAAGTCCCGAAGAGATCTCATCGTCAT 479
Qy 290 t-----GluIleAsnProLysAsp-----ThrValLeuThrLeuTh 302
Db 480 GCTCAGGACACTGACCTGACGTAACAAGAGGACAAAGAGGACTGCTCTACATCTGGC 539
Qy 302 rSer-----GlyGlyCysAsnAlaLeuAsnLeuValGlnGlyAlaGlyGlnValVa 320
Db 540 CTCTGCCAATGGGAATTCAGAAAGTAGTAAACCTCTCTGGACAGACGATGTCAACTTAA 599
Qy 320 lSerValAspCysAsnProAlaGlnSerAlaLeuLeuGluLeuLysLysValAlaIleGl 340
Db 600 TGTCCTTGAC---AACAAGAGGACAGCTCTGATAAG-----GCCGTACA 644
Qy 340 nGlnLeuGluPheGluAspValTrpGlnLeuPheGlyGluGlyValHisProArgIleGl 360
Db 645 ATGCCAGGAGATGAATGCTGTTAATGTTGCTGGACATGGCACTGATCAATAATTC 704
Qy 360 uGluLeuTyrl-GluLysLysLeuAlaProPheLeuSerGlnThrSerHisAsnPheTrpS 380
Db 705 AGATGAGTATGGAATACCACTCTGCCTACTAGCTATCTATAATGAAGATAAATTAATGGC 764
```

Qy 380 erLysArgLeuTrpTyr 385
Db 765 CAAGCACTGCTCTTAT 781

Search completed: March 14, 2006, 07:38:00
Job time : 324 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 14, 2006, 05:00:51 ; Search time 1369 Seconds
(without alignments)
3914.213 Million cell updates/sec

Title: US-10-620-914-45

Perfect score: 3463

Sequence: 1 MGSGRDRPASVTKNFSLE.....RVNYSFFYMARKGAKKON 648

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DRV=xlh
-O=/abs/ABSSWEB.spool/US10620914/runat.13032006.102049.6893/app_query.fasta_1
-DB=Published Applications_NA_Main -QFWT=fastap -SUFFIX=p2n.rnpbm
-MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1
-MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -MODE=LOCAL -OUTFMT=ptc -NORM=ext
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -DOCL=LOCAL -OUTFMT=ptc -NORM=ext
-HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abss05h
-USER=US10620914 @CGN 1.1.1026 @runat.13032006.102049.6893 -NCPU=6 -ICPU=3
-NO_WMAP -NEG_SCORES=0 -WAIT -DSEBLOCK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPEXT=0.5 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEX=7

Database : Published Applications_NA_Main:

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3463	100.0	1947	7	US-10-620-914-44
2	2838	82.0	5242	7	US-10-620-914-43
3	891	25.7	2259	8	US-10-741-849-6021
4	850.5	24.6	2727	7	US-10-620-914-49
5	804	23.2	3427	7	US-10-620-914-48
6	379.5	11.0	1252	5	US-10-118-495-1
7	379.5	11.0	1252	7	US-10-620-914-1

8	329	9.5	1248	5	US-10-118-495-22	Sequence 22, Appl
9	329	9.5	1248	7	US-10-620-914-22	Sequence 22, Appl
10	326.5	9.4	1251	5	US-10-118-495-28	Sequence 28, Appl
11	326.5	9.4	1251	7	US-10-620-914-28	Sequence 28, Appl
12	292.5	8.4	1251	5	US-10-118-495-32	Sequence 32, Appl
13	292.5	8.4	1251	7	US-10-620-914-32	Sequence 32, Appl
14	128	3.7	4246	6	US-10-085-959-12	Sequence 12, Appl
15	123	3.6	7119	8	US-10-684-141-57	Sequence 57, Appl
16	123	3.6	7119	9	US-10-810-486-57	Sequence 57, Appl
17	120	3.5	666	5	US-10-118-495-34	Sequence 34, Appl
18	120	3.5	666	7	US-10-620-914-34	Sequence 34, Appl
19	119	3.4	6912	8	US-10-335-053-309	Sequence 309, Appl
20	119	3.4	6912	8	US-10-684-141-53	Sequence 53, Appl
21	119	3.4	6912	9	US-10-810-486-53	Sequence 53, Appl
22	119	3.4	6912	9	US-10-887-553A-400	Sequence 400, Appl
23	118.5	3.4	1155	3	US-09-825-301-5	Sequence 5, Appl
24	118.5	3.4	1155	3	US-09-759-143-373	Sequence 373, Appl
25	118.5	3.4	1155	3	US-09-780-669-373	Sequence 373, Appl
26	118.5	3.4	1155	3	US-09-810-936-301	Sequence 301, Appl
27	118.5	3.4	1155	3	US-09-822-827-373	Sequence 373, Appl
28	118.5	3.4	1155	3	US-09-429-755-301	Sequence 301, Appl
29	118.5	3.4	1155	3	US-09-924-400-301	Sequence 301, Appl
30	118.5	3.4	1155	3	US-09-895-793-373	Sequence 373, Appl
31	118.5	3.4	1155	3	US-09-895-814-373	Sequence 373, Appl
32	118.5	3.4	1155	5	US-10-012-896-373	Sequence 373, Appl
33	118.5	3.4	1155	5	US-10-010-940-373	Sequence 373, Appl
34	118.5	3.4	1155	6	US-10-212-679-301	Sequence 301, Appl
35	118.5	3.4	1155	6	US-10-144-678A-373	Sequence 373, Appl
36	118.5	3.4	1155	6	US-10-033-527-5	Sequence 5, Appl
37	118.5	3.4	1155	6	US-10-294-025-373	Sequence 373, Appl
38	118.5	3.4	1155	7	US-10-079-137B-301	Sequence 301, Appl
39	118.5	3.4	1185	3	US-09-924-400-335	Sequence 335, Appl
40	118.5	3.4	1185	6	US-10-212-679-335	Sequence 335, Appl
41	118.5	3.4	1185	7	US-10-079-137B-335	Sequence 335, Appl
42	118.5	3.4	1383	3	US-09-905-673-64	Sequence 64, Appl
43	118.5	3.4	1383	6	US-10-096-319-64	Sequence 64, Appl
44	118.5	3.4	1434	3	US-09-095-673-65	Sequence 65, Appl
45	118.5	3.4	1434	6	US-10-096-319-65	Sequence 65, Appl

ALIGNMENTS

RESULT 1
US-10-620-914-44
; Sequence 44, Application US/10620914
; Publication No. US20040093639A1
; GENERAL INFORMATION:
; APPLICANT: Benning, Christoph
; APPLICANT: Riekhof, Wayne
; TITLE OF INVENTION: Klug, Rouven
; FILE REFERENCE: MSU-07769
; CURRENT APPLICATION NUMBER: US/10/620,914
; PRIOR FILING DATE: 2003-07-16
; PRIOR FILING DATE: 2002-04-08
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patent version 3.2
; SEQ ID NO 44
; TYPE: DNA
; LENGTH: 1947
; ORGANISM: Chlamydomonas reinhardtii
US-10-620-914-44

Alignment Scores:	0	Length:	1947
Pred. No.:	3463.00	Matches:	648
Score:	100.0%	Conservative:	0
Best Local Similarity:	100.0%	Mismatches:	0
Query Match:	100.0%	Indels:	0
DB:	7	Gaps:	0
US-10-620-914-45 (1-648) x US-10-620-914-44 (1-1947)			

QY 1 MetGlySerGlyArgAspGlyArgProAlaSerTyrThrLysLysAsnPheSerLeuGlu 20
DB 1 ATGGGGTGGGTCTGACGGCGGCTCGAGCTACACCAAGAGAATCTTCTCCCTGGAG 60
QY 21 LysLeuLysLeuSerSerMetLysAspAspLeuThrValLeuArgHisMetTrpPheGly 40
DB 61 AAGCTCAAGCTCAGCAGCATGAAGGATGACCTGACCGTCTCTGGCCATATGTGTTCGGC 120
QY 41 SerLysLysGlyAspAspHisAlaAlaArgLeuGluSerPheTyrGlyProGlnAlaAla 60
DB 121 AGCAAGAAGCGGATGATCATCGCTGCTCGCTGAGAGCTTCTACGGGCGCCCAAGCGCGCT 180
QY 61 AlaPheAlaAlaArgLeuAlaGluArgSerAsnLeuLeuTrpValAspLeuGlyGly 80
DB 181 GCCTTTCTGCCCGCTGGCGGAGCGCTCGAACCTCATCTGGGTGTGACCTGGGTGGTGGC 240
QY 81 ThrGlyGluAsnValAspMetMetAlaAspTyrLysLeuAlaLysPheLysSerIle 100
DB 241 ACTGGGAGAAATGTCGATATGATGCTGATTTACATCGACCTGGCGAAGTTTCAAGTCCATC 300
QY 101 TyrValValAspLeuCysHisSerLeuCysGluValAlaLysLysLysAlaLys 120
DB 301 TACGTGTCTGACCTGTGCCATCTGCTGTGGAGGTGGCCCAAGAAGAAGCGCAAG 360
QY 121 GlyTrpLysAsnValGlnValValGluAlaAspAlaCysGlnPheAlaProProGluGly 140
DB 361 GGCTGGGAAGAATGTCAGGTGCTGGAGGCGACCTTGGCAATTTGGCGCCCTTGAGGGC 420
QY 141 ThrAlaThrLeuIleThrPheSerTyrSerLeuThrMetIleProProPheHisAsnVal 160
DB 421 ACCGCGACGCTCATCCTCTCTCTACTCGCTCAGCATGATTCCACCGTTCCCAACGTC 480
QY 161 IleAspGlnAlaCysSerTyrLeuSerGlnAspGlyLeuValGlyValAlaAlaAspPheTyr 180
DB 481 ATCGACGAGGCTTCTCGTACCTGCTCCCAAGACGGCTGTGGTGGCGCTTCTCTAC 540
QY 181 ValSerGlyLysTyrAspLeuProLeuArgGlnMetProTrpSerArgArgPhePheTyr 200
DB 541 GTGAGCGGCAAGTACGACCTGCTCCCTCGCCAGATGCTGCTGGTGGCGCTTCTCTGG 600
QY 201 ArgSerIlePheAspIleAspAsnIleAspIleGlyProGluArgArgAlaTyrLeuGlu 220
DB 601 CGATCGATCTTCGACATCGACCAATGACATCGGCGCCCGAGCGCGCTACCTGGAG 660
QY 221 GlnLysLeuGluArgValTrpGluGlnAsnThrGlnGlySerIleProTyrValProTrp 240
DB 661 CAGAAGCTGGAGCGCGTGTGGGAGCAGAACACCCAGGGTTCGATCCCTACGTGCCGTGG 720
QY 241 LeuArgAlaProTyrTyrValTrpIleGlyArgLeuProSerValGlyHisAlaLeuHis 260
DB 721 CTGGCGCCCTTACTACGTGTGATTTGGCGCTGCTCCAGCGTGGCCAGCGCTCGAC 780
QY 261 GluGluArgValGluArgProProMetPheProProThrPheLeuTyrThrGlnSerTrp 280
DB 781 GAGGAGCGGTGGAGCGCGCCCATGTTCGCGCCACCTTCTGTACACGAGTCTGTGG 840
QY 281 GluAspProGluProAspMetGluValMetGluIleAsnProLysAspThrValLeuThr 300
DB 841 GAGGACCCCGAGCGGATATGAGGTATGAGATCAACCCCAAGAGACACGCGTCTGACC 900
QY 301 LeuThrSerGlyGlyCysAsnAlaLeuAsnLeuValGlnGlyAlaGlyGlnValVal 320
DB 901 CTGACTAGCGCGCTGCATGCTCCCTGACCTGTGTGGTGGAGGGCGCGCAGGTGGTG 960
QY 321 SerValAspCysAsnProAlaGlnSerAlaLeuLeuGluLeuLysLysValAlaIleGln 340
DB 961 TCGGTGGACTGCAACCCCGCGAGTCGCGCTTCTGGAGCTGAAGAAGGTGGCCATTCAG 1020
QY 341 GlnLeuGluPheGluAspValTrpGlnLeuPheGlyGluGlyValHisProArgIleGlu 360
DB 1021 CAGCTGAGTTTGGAGCGTGTGGCAGCTGTTCGGGAGGGCGTGGCAGCGCATTTGAG 1080

QY 361 GluLeuTyrGluLysLysLeuAlaProPheLeuSerGlnThrSerHisAsnPheTrpSer 380
DB 1081 GACCTGTACGAGAAGAAGCTGGCGCTTCTCTGTCAAACACGACCACTTCTGTGTCC 1140
QY 381 LysArgLeuTrpTyrPheGlnHisGlyLeuTyrTyrGlnGlyMetGlyLysLeuCys 400
DB 1141 AAGCGCTCTGTGTACTTCCAGCAGCGCTGTACTACAGCGCGCATGGCAAGTGTGC 1200
QY 401 TrpValLeuGlnCysLeuAlaValValLeuGlyLeuGlyLysThrValLysArgLeuAla 420
DB 1201 TGGTGTCTGAGTGGCTTGGCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGC 1260
QY 421 AsnAlaProThrMetGluGluGlnArgArgLeuTrpAspSerAsnMetLeuIleHisPhe 440
DB 1261 AACGCGCCCAACATGGAGGAGCAGCGCTGTGTGGGACAGCAACATGCTCATCCACTTC 1320
QY 441 ValLysAsnGlyProLysProLeuValTrpLeuPheValLysPheValSerLeuValLeu 460
DB 1321 GTGAAGAACCGGCGCAAGCGCTGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGC 1380
QY 461 PheAsnLysAlaValLeuTrpPheGlyGlyValProGlyLysGlnTyrAlaLeuIle 480
DB 1381 TTCAACAAGCCGTGTGTGTGGCGGGCGGTGGCGGGCAAGCAAGTACGGCTGATC 1440
QY 481 LysAlaAspGlyIleProLysGluAsnTyrIleAlaArgThrMetAspGlyValAlaGlu 500
DB 1441 AAGCGGAGCGCATCCCATTTGAGAACTACATCGCGCACCATGGAGCGCGTGGCGGAG 1500
QY 501 AsnSerHisValArgLysGlnAsnTyrPheTyrTyrAsnCysLeuThrGlyLysPheLeu 520
DB 1501 AACTCGCAGCTGGCGAAGCAGAACTACTTCTACTACAACCTGCCTCACCGCAAGTCTCTG 1560
QY 521 ArgAspAsnCysProThrTyrLeuArgGluAlaAlaPheAlaThrLeuLysSerGlyVal 540
DB 1561 CGGCAACAACCTGCCCACTTACCTTGGCGAGGGCGCTTCCCACTCCCAAGAGTGGCGTG 1620
QY 541 ValAspAsnLeuThrValSerThrAsnPhePheMetGluLeuLysAlaArgThrTyr 560
DB 1621 GTGGACAACCTGACCGCTCTCCACCACTTCTTATGGAGGAGCTCAAAGCGCGCACCTAC 1680
QY 561 ThrLysValIleLeuMetAspHisValAspTrpLeuAspMetProValAlaAsnGluLeu 580
DB 1681 ACCAAGTGTATTTGTATGGACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1740
QY 581 AlaGluCysLeuAlaLysGlnValAlaProGlyGlyValIleTrpArgSerAlaSer 600
DB 1741 GCGAGTGGCTGGCCAAAGCAGGTGTGGCGGGCGGCATCTCATCTGGCGCTCGCCCTCC 1800
QY 601 LeuSerProProTyrAlaGluLeuIleGlnLysAlaGlyPheAspValArgCysIleArg 620
DB 1801 CTGAGCGCGCTTACGCGAGCTGATCCAGAAGCGGGCTTCGACGTGGCTGCATCCGC 1860
QY 621 ArgAlaThrGlnGlyTyrMetAspArgValAsnMetTyrSerSerPheTyrMetAlaArg 640
DB 1861 CGCGCCACTCAGGGCTACATGGACCGGTCAACATGTACAGCTCTCTTCTCATGSGCCGC 1920
QY 641 ArgGlyGlyAlaLysLysAspAsn 648
DB 1921 CGGAAGGGCGCCCAAGAAGGACAAC 1944

RESULT 2

US-10-620-914-43
; Sequence 43, Application US/10620914
; Publication No. US20040093639A1
; GENERAL INFORMATION:
; APPLICANT: Benning, Christoph
; APPLICANT: Riekhof, Wayne
; APPLICANT: Klug, Rouven
; TITLE OF INVENTION: Compositions and Methods for the Production of Betaine Lipids
; FILE REFERENCE: WSU-0769
; CURRENT APPLICATION NUMBER: US/10/620,914
; CURRENT FILING DATE: 2003-07-16
; PRIOR APPLICATION NUMBER: 10/118,495

Qy 288 ----- 288
Db 2418 GGCAGATACGACGAGAGAAACGGTCTGTGTACCCCGAGGGCGGGATCACGACGTTCCG 2477
Qy 288 ----- 288
Db 2478 AAGCATTGTGGCAGGTACATCGTCACCAGCAAGCAAGCACTAAGCACCAGCGGTT 2537
Qy 288 ----- 288
Db 2538 CTGTGGCAATGTGCCACCGCAATGCTGGCCAGTGGCAGCTTCGCAATTAGTGTATGCC 2597
Qy 288 ----- 288
Db 2598 AGTATCAGCGCTAGCTCAGCGCTGCAAGCTGCTGTGTAGAAAGCAGCGGATGGTGGCACT 2657
Qy 289 -----ValMe 290
Db 2658 GAGCGGAACTCTCTGAACTGTGCTGCCATCTCTCTGTCCTTCCCAGGTGAT 2717
Qy 290 tGluIleAsnProIysAspThrValLeuThrLeuThrSerGlyGlyCysAsnAlaLeuAs 310
Db 2718 GGAGATCAACCCCAAGACACGGTGTGACCTGACTAGCGCGGCTGCAATGCCCTGAA 2777
Qy 310 nLeuLeuValGlnGlyAlaGly ----- 317
Db 2778 CTTGCTGGTGAGGGGCCGGCCAGGTGAGATCATAGTGGTGTGACACCTTTCTGGTG 2837
Qy 317 ----- 317
Db 2838 TCCTTACTTCATGATGCGCTGACGGAATGTTAGGAAGCGGTGGCATACATGTAGCT 2897
Qy 318 -----GlnValValSerValAspCysA 325
Db 2898 GCAGAAATGCTGACGTGCTTTCCGCGCTGTGCTGTCAGGTGGTGTGCTGGTACTGCA 2957
Qy 325 snProAlaGlnSerAlaLeuLeuGluLeuLysValAlaLleGlnGlnLeuGluPheG 345
Db 2958 ACCCGCGCAGTCCGCGCTCTTGAGCTGAAGAAGGTGGCCATTTCAGCAGCTGGAGTTG 3017
Qy 345 luAspValTtp ----- 348
Db 3018 AGGAGGTGTGGCAGGTAAAGGGCTCCTCATCTGCGCGCTAGGACTGCACAGCTCTGTT 3077
Qy 348 ----- 348
Db 3078 CTACTGTGATGCAACATGCGGGACCTGGTCAGGCTTTCGCGCAFTCGCACTGCGATGCTG 3137
Qy 348 ----- 348
Db 3138 TCCCCAGGGCTCAGCAATCTCTCGCGGCTGCAACGGCAGCGCCCTTTTACACTCC 3197
Qy 349 -----GlnLeuPheGlyGluGlyValHisProArg 358
Db 3198 GCRAACCTCTGTTGGTAATTCCTTTCCACAGCTGTTGCGGAGGGCGTGCAACCGCGC 3257
Qy 359 IleGluGluLeuTyrrGluLysLysLeuAlaProPheLeuSerGlnThrSerHisAsnPhe 378
Db 3258 ATTGAGGAGCTGTACGAGAAGAAGCTGGCGCTTCTCTGCGCAACAGCAGCAACTTC 3317
Qy 379 TrpSerLysArgLeuTtpTyrrPheGlnHisGlyLeuTyrrGlnGlyGlyMetGlyLys 398
Db 3318 TGGTCCAAAGCCCTCTGTGTACTTCAGACAGGCTGTACTACAGGCGGATGGGCAAG 3377
Qy 399 LeuCysTrpValLeuGlnCysLeuAlaValValLeuGlyLeuGlyLysThrValLysArg 418
Db 3378 CTGTGCTGGTGTGTCAGTGTGCTGGCTGGTGTGGTGTGGCAAGCGTCAAGCGC 3437
Qy 419 LeuAlaAsnAlaProThrMetGluGlnArgArgLeuTrpAspSerAsnMetLeuIle 438
Db 3438 CTCGCCAACCGCCCAATGGAGGAGCGCGCTGTGTGGGACAGCAATCTGTCATCTCATC 3497

Qy 439 HisPheValLysAsnGlyProLysProLeuValTrpLeuPheValLysPheValSerLeu 458
Db 3498 CACTTCGTGAAGAACGGCCCAAGCGCTGGTGTGCTGTTCGTCAAGTTCGTGAGCGTG 3557
Qy 459 ValLeuPheAsnLysAlaValLeuTrpPheGlyGlyValProGlyLysGlnTyrrAla 478
Db 3558 GTGCTCTTCAACAAAGCCGTGTGTGTTCCGGCGGGCGGTGCGCGCAAGCAAGTACGGC 3617
Qy 479 LeuIleLysAlaAspGlyIleProIleGluAsnTyrrIleAlaArgThrMetAspGlyVal 498
Db 3618 CTGATCAAGCGGACGGCATCCCATTTGAGAACTACATCGCGCCACCATTGGACGGCGTG 3677
Qy 499 AlaGluAsnSerHisValArgLysGlnAsnTyrrPheTyrrAsnCysLeuThrGlyLys 518
Db 3678 GCGGAGAACTCGACCTGCGCAAGCAGAACTACTTCTACTACAACTGCCTCACCGCAAG 3737
Qy 519 PheLeuArgAspAsnCysProThrTyrrLeuArgGluAlaPheAlaThrLeuLysSer 538
Db 3738 TTCTCTCGGACAACTGCCCCACCTACCTGCGGAGCGGCTTCGCCACCTCAAGAGT 3797
Qy 539 GlyValValAspAsnLeuThrValSerThrAsnPheMetGluGluLeuLysAlaArg 558
Db 3798 GCGGTGGTGACAACTGACCGTCTCCACCACTTCTTCATGAGGAGCTCAAAGCGCGC 3857
Qy 559 ThrTyrrThrLys ----- 562
Db 3858 ACCTACACCAAGTGGGAGCCAGCGAGAGTGTGAAAAGCGGAGATGGCGGCGTTG 3917
Qy 562 ----- 562
Db 3918 TAGGTGATGCGCAACGCTTGTCTGCCAGGTTGAAAGGCGAGAGCGAGGGAACCGAATT 3977
Qy 562 ----- 562
Db 3978 TCAGAAATTCGATGGCTGGGAACCGGAACCAAGAACGGAATAGAAAGCGCTGGCG 4037
Qy 562 ----- 562
Db 4038 AGAACTGGTGGCTTGGCGTTTGGCATCCAGTCTGCCAGCGCGCCCGCTACCTGCT 4097
Qy 563 -----Val 563
Db 4098 CAGGTCCGTCTAACCGTTCAACCCCTCGCCCTCACCGCCCCCCCCCGCCACCGAGTG 4157
Qy 564 IleLeuMetAspHisValAspTrpLeuAspMetProValAlaAsnGluLeuAlaGluCys 583
Db 4158 ATTCTGATGGACCACCGTGGACTGGCTGGATATGCCGTGGCCCAACGAGCTGGCGAGTGC 4217
Qy 584 LeuAlaLysGlnValAlaProGlyGlyIleValIleTrpArgSerAlaSerLeuSerPro 603
Db 4218 CTGGCAAGCAGGTTGGCGCGGCGGCATCGTCATCTGGCGCTCCGCTCCCTCAGCGCG 4277
Qy 604 ProTyrrAlaGluLeuIleGlnLysAlaGlyPheAspValArgCysIleArgAlaThr 623
Db 4278 CCCTACGCCGAGCTGATCCAGAAAGCGGGCTTCGACGTGCGTGCATCCCGCGCGCACT 4337
Qy 624 GlnGlyTyrrMetAspArgValAsnMetTyrr ----- 633
Db 4338 CAGGCTTACATGACCGCGTCAACATGTACAGGTGCGTGTCTTGTCCGTGAGGTGTGAG 4397
Qy 633 ----- 633
Db 4398 GATCTAGGATGTCAGTGGTGGGAGGAGCATGCGGGTCAATGGGAGGAGGTTCGCT 4457
Qy 633 ----- 633
Db 4458 TATTGATGCCGATGGAGCCTGGGTGAATCGGTGAGCGGCTACCGTATGTGCCAAGGAC 4517
Qy 633 ----- 633
Db 4518 GGGGTTGGCGGCTCGGGCGGACAGAAACTTGAAGGTTGTTTTAGAAAGATCAAGATA 4577
Qy 633 ----- 633

Db 4578 TGTTCGATACGAGGACGCTAGCGTCAGGGCGCAACAGCGCGTGGCGCACGCGGAAA 4637
 QY 634 -----Ser 634
 Db 4638 TCCTGGTTCCTTCGATGATATACCTGATCCAACTCTTGTGGTTCTCCATGGCGAGC 4697
 QY 635 SerPheTyrMetAlaArgArgGlyGlyAlaLysAlaAspAsn 648
 Db 4698 TCCTTCTACATGCGCGCGGAGGCGCCCAAGAGGACAAAC 4739

RESULT 3

US-10-741-849-6021
 ; Sequence 6021, Application US/10741849
 ; Publication No. US20050019931A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Roemer, Terry
 ; APPLICANT: Jiang, Bo
 ; APPLICANT: Boone, Charles
 ; APPLICANT: Bussey, Howard
 ; TITLE OF INVENTION: Nucleic Acids Encoding Anti-fungal Drug Targets and Methods of
 ; FILE REFERENCE: 10182-023-999
 ; CURRENT APPLICATION NUMBER: US/10741,849
 ; CURRENT FILING DATE: 2003-12-19
 ; PRIOR APPLICATION NUMBER: US 60/434,832
 ; PRIOR FILING DATE: 2002-12-19
 ; NUMBER OF SEQ ID NOS: 8000
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 6021
 ; LENGTH: 2259
 ; TYPE: DNA
 ; ORGANISM: Candida albicans
 US-10-741-849-6021

Alignment Scores:

Pred. No.: 3,81e-99 Length: 2259
 Score: 891.00 Matches: 229
 Percent Similarity: 46.6% Conservative: 98
 Best Local Similarity: 32.7% Mismatches: 232
 Query Match: 25.7% Indels: 142
 DB: 8 Gaps: 21

US-10-620-914-45 (1-648) x US-10-741-849-6021 (1-2259)

QY 51 LeuGluSerPheTyrGlyProGlnAlaAlaAlaPhe----- 62
 Db 208 TTAGAGTCATCTACAGAAATCAAGCTCATATTTATGATACACTAGAGAAATTTTGTG 267
 QY 63 -----AlaAlaArgLeuAlaGluArgSerAsnLeu 72
 Db 268 AAAGGTAGACAAGATGCTTAGATTAGCTATTTCCATTTACCCAAAAGAAAGATCTT 327
 QY 73 IleTyrValAspLeuGlyGlyThrGlyGluAsnValAspMetMetAlaAspTyrIle 92
 Db 328 ATTTGGATGATATGCTGGTGAACCTTGGTTCCAAATATTGAATTCATGGATGAATTAGT 387
 QY 93 AspLeuAlaLys---PheLysSerIleTyrValValAspLeuCysHisSerLeuGlu 111
 Db 388 AAAATATCTGAAAACCTTTAAAGCTGTTATTTGGTTGATCTTTCCCATCTTTTGTGTGAA 447
 QY 112 ValAlaLysLysLysAlaLysAlaLysGlyTyrLysAsnValGlnValValGluAlaAsp 131
 Db 448 GTTGCTAAGCAGAGATTGAGCCCATGATGATGACAAAGTTTCATGATTTAGTTGCTGAT 507
 QY 132 AlaCysGlnPheAlaProGluGlyThrAlaThrLeuIleThrPheSerTyrSerLeu 151
 Db 508 GCCTGTGATTTACTATTGATTATGATGCTGCTGAATTCATTTCTTATTCTATTG 567
 QY 152 ThrMetIleProPheHisAsnValIleAspGlnAlaCysSerTyrLeuSerGlnAsp 171
 Db 568 TCGATGATCCAACTTTCAATGCTGCTATCGAATGCTGTTCTTCAATTAGATATGGAA 627

QY 172 GlyLeuValGlyValAlaAspPheTyrValSer-----GlyLysTyr 185
 Db 628 GGTATTATTATTCACCTGCTGGATTTGGTATTCAAAGCAGTGACACCTCAATTCGTCGATC 687
 QY 186 Asp-----LeuProLeuArgGlnMetProTyrSerArgPhePheTyrArg 201
 Db 688 AATACCTGTTGGTGGTGGTTTACAGGACATTCCTTGGATATTATGTAATTTTGGAGA 747
 QY 202 SerIlePheAspIleAspAsnIleAspIleGlyProGluArgArgAlaTyrLeuGluGln 221
 Db 748 ATTTGGTTTGAAGCTGATAAAGTGTGTTTGGATCTTCAAGAAGAAACATTTTGAATAT 807
 QY 222 LysLeuGluArgValTyrGluGlnAsnThrGln-----GlySerIleProTyr 237
 Db 808 AAATTTGGTACCGTCAAAATCTTTGAATTCATACAACAAGGCTTTGGGTAAATC----- 861
 QY 238 ValProTyrLeuArgAlaProTyrTyrValTyrIleGlyArgLeuProSerValGlyHis 257
 Db 862 -----CCCTATTATATTGATTTGGATTTGGTATGAATTAATCAAAATCAAC 903
 QY 258 AlaLeuHisGluGlu-----ArgValGluArgProProMetPheProProThr 273
 Db 904 ACCATTTTGAAGAGATTGAATTTGTAGCCACTGATCCCTTACCTTGCTCCCACTACA 963
 QY 273 ----- 273
 Db 964 ACTCCAATCGCTAATCAACTTGAAGATATTCCAATTTCTAAAGGTCATGAAGCTGCTTTA 1023
 QY 273 ----- 273
 Db 1024 ATCAACTTGCAAAAAATTTACCTTACCCTCAATCTACTATCAAAAGGAATATTGGAGA 1083
 QY 274 -----PheLeuTyrThr 277
 Db 1084 GTCTACTATGATGAATGAATCCATTGTATGAACATTTTAAACCAATACATATTATGCT 1143
 QY 278 GlnSerTyrGluAspProGluProAspMetGluValMetGluLeuAsnProLysAspThr 297
 Db 1144 TTCACCTTGGAAAGATCTCGTGAAGATCATAACTTTTGAATTTTACCAGTGTATGATCT 1203
 QY 298 ValLeuThrLeuThrSerGlyGlyCysAsnAlaLeuAsnLeu-----LeuValGlnGly 315
 Db 1204 GTTTTGGCTATTACTTTCAGCTGGTGATATAATTTTGAAGTATGCTAGTTTACCACACCA 1263
 QY 316 AlaGlyGlnValValSerValAspCysAsnProAlaGlnSerAlaLeuLeuGluLeuLys 335
 Db 1264 CCAAAAAGAGATCATCTGCTGATCTTATCCATGTCATAAACCATTTATTAGATTTGAAA 1323
 QY 336 LysValAlaIleGlnGlnLeuGluPheGluAspValTyrGlnLeuPheGlyGluGlyVal 355
 Db 1324 TTGGCTAGTTTGTAGATGCTCTTCTCAAGAACAAATTTGGTCAATGTTTGGTGAAGGTAA 1383
 QY 356 HisProArgIleGluGluLeuTyrGluLysLeuAlaProPheLeuSerGlnThrSer 375
 Db 1384 ATCGAAAAATTCATGATCTTTTGTATGATACCTTTGGCCCGCACATGCTTCTAATGCC 1443
 QY 376 HisAsnPheTyrSerLysArg-----LeuTyrTyrPheGlnHisGlyLeuTyrGln 393
 Db 1444 TTCCAATATGATGATGATGAAGACCTAAACCTTTTCTGGTAAAGGTCTTTAT----- 1497
 QY 394 GlyGlyMetGlyLysLeuCysTyrValLeuGlnCys-----LeuAlaValValLeu 410
 Db 1498 ---GATACTGGGTTTCTAGATGGGCATTTAAGATTTATCAAGATATGTTTCAAAAGTTGC 1554
 QY 411 GlyLeuGlyLysThrValLysArgLeuAlaAsnAlaProThrMetGluGlnGlnArgArg 430
 Db 1555 GGTGTGTAGTAAATACCTTGAAGAACTTTGTGCTGCAACCCACCATGAAGAACTATTGGA 1614
 QY 431 LeuTyrAspSerAsnMetLeuIleHisPheValLysAsnGlyProLysProLeuValTyr 450
 Db 1615 ATTTGAATGAACATTTG-----AAACCACT----- 1641
 QY 451 LeuPheValLysPheVal---SerLeuValLeuPheAsnLysAlaValLeuTyrPheGly 469

```
Db 1642 TTAATTAATCTGTGGTGGTTCATTACTTGTGGTAAATCAATGTTTATGGAAGCT 1701
Qy 470 GlyGlyValProGlyLysGlnTyrAlaLeuLeuLysAlaAspGlyIleProIleGluAsn 489
Db 1702 TTGGGAGTTCAGCTAATCAAGCAGCTTTAATG-----GGACCATCTGTATTAAA 1752
Qy 490 TyrIleAlaArgThrMetAspGlyValAlaGluAsnSerHisValArgLysGlnAsnTyr 509
Db 1753 TAGTGTGTTGATCTTTGGGCCCATCATTAAGAGATCGATGATTTCAATGATTAATAT 1812
Qy 510 PheTyrTyrAsnCysLeuThrGlyLysPheLeuArgAspAsnCysProThrTyrLeuArg 529
Db 1813 TTCTACTATTATGATATGATGGGAGATACACCAAAACAATGTCAGATTATTTAACT 1872
Qy 530 GluAlaAlaPheAlaThrLeuLys-----SerGlyValVal 541
Db 1873 ACAAAGGTTTCAACAGATTATCTAGCACTGCTGCTACTGCCAGCGGATCATCTCCAATT 1932
Qy 542 AspAsnLeuThrValSerThrAsnPhePheMetGlu-----GluLeuLysAlaArg 558
Db 1933 GACAACTTGAATCCACACTGACACTTTAAATGAAGTGTGGTAGATTAAAGAAAAA 1992
Qy 559 ThrTyrThrLysValIleLeuMetAspHisValAspTrpLeuAsp-----MetPro 575
Db 1993 TCAATCACTATGCAATATCATGATCATATGATGTTTGAACCCCTAATGGTAGAGAT 2052
Qy 576 ValAlaAsnGluLeuAlaGluCysLeuAlaLysGlnValAlaProGlyGlyIleValIle 595
Db 2053 GCTATTAATGAATTAATCT---GCTTTGAAAAGATGCTTGTCCCGAGGGGTAGATATTA 2109
Qy 596 TrpArgSerAlaSerLeuSerProTyrAlaGluLeuIleGlnLysAlaGlyPhe--- 614
Db 2110 CTTAGATCAGCAAGTCAAGACCTTGTGACTTGAACATTCAGAACTTGGGATCCAA 2169
Qy 615 AspValArgCysIleArgAlaAlaThrGlnGlyTyrMetAspArgValAsnMetTyrSer 634
Db 2170 GAAGAAGAAAATGTTGTTGCTGCAACCTGGTTCAAGTATAGACAGAGTTAATATGATGCC 2229
Qy 635 Ser 635
Db 2230 AAT 2232
```

RESULT 4

```
US-10-620-914-49
; Sequence 49, Application US/10620914
; Publication No. US20040093639A1
; GENERAL INFORMATION:
; APPLICANT: Benning, Christoph
; APPLICANT: Riekhof, Wayne
; APPLICANT: Klug, Rouven
; TITLE OF INVENTION: Compositions and Methods for the Production of Betaine Lipids
; FILE REFERENCE: MSU-07769
; CURRENT APPLICATION NUMBER: US/10/620,914
; CURRENT FILING DATE: 2003-07-16
; PRIOR APPLICATION NUMBER: 10/118,495
; PRIOR FILING DATE: 2002-04-08
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 49
; LENGTH: 2727
; TYPE: DNA
; ORGANISM: Neurospora crassa
US-10-620-914-49
```

```
Alignment Scores:
Pred. No.: 5,71e-94 Length: 2727
Score: 850.50 Matches: 239
Percent Similarity: 44.4% Conservative: 109
Best Local Similarity: 20.5% Mismatches: 240
Query Match: 24.6% Indels: 198
DB: 7 Gaps: 23
```

```
US-10-620-914-45 (1-648) x US-10-620-914-49 (1-2727)
Qy 19 LeuGluLysLeuLysLeuSerSerMetLysAspLeuThrValLeuArgHisMet--- 37
Db 449 TTGAGAGACTGGAGATG---GGCTTGAANAAGACAAGAGATGGGAACATGTGTAA 505
Qy 38 ---TrpPheGlySerLysLysGlyAspAspHisAlaAlaArgLeuGluSerPheTyrGly 56
Db 506 CCGTGGCGGACAG-----GGA 523
Qy 57 ProGlnAlaAlaPheAlaAlaArgLeuAlaGluArgSerAsnLeuIleTrpValAsp 76
Db 524 CCG-----GGCAAGGAGGAACCG-ATATGGGTAGAT 555
Qy 77 LeuGlyGlyGlyThrGlyGluAsnValAspMetAlaAspTyrIleAspLeuAlaLys 96
Db 556 GTCGGTGGGGGCACAGCTGGATATATCGAAGCCATGGCCAAGTTTGTCAACGCTCTGAA 615
Qy 97 ---PheLysSerIleTyrValValAspLeuCysHisSerLeuCysGluValAlaLysLys 115
Db 616 TTCTTCAAGACTGTTTACCTAGTGGACTTTTCGCCGTCACCTTTGTGAAGTGGCTAGGAAG 675
Qy 116 LysAlaLysAlaLysGlyTrpLysAsnValGlnValValGluAlaAspAlaCysGlnPhe 135
Db 676 CGGTTTGGCCAGGCTGGGGTGGGAGAATGTGAGAGTTATCTGCACGGATGTCGCAAGTTT 735
Qy 136 Ala-----ProProGluGlyThr----- 141
Db 736 AGGCTTGAGGATTTATGAGGATGTTGAGAAGAGAGAGTCTGGCTCTGGAGATTTCTGCCT 795
Qy 142 -----AlaThrLeuIle 145
Db 796 TCTTTGTGGGTTGGTGGGGGAGACGAAGCCGGACACATCGGGAGCTGAGTTGATC 855
Qy 146 ThrPheSerTyrSerLeuThrMetIleProPheHisAsnValIleAspGlnAlaCys 165
Db 856 ACCATGCTTATAGACCTTTCGATGATGCCGATTTATTTCTCGATTATCGATTCTCGTCGAG 915
Qy 166 SerTyrLeuSerGlnAspGlyLeuValGlyValAlaAspPheTyrValSerGlyLysTyr 185
Db 916 TCTCTGTAGACCTCACGCTTGATGTCGCTGTGGACTTTTACGCCCATGCGAAGTC 975
Qy 186 Asp-----LeuProLeuArgGlnMetProTyrSerArg 196
Db 976 GACTTCACATTCGCCAATACACGGGTGGTCTTATGAACCGACACACGTGTGGCTATTTCCGC 1035
Qy 197 ArgPhePheTrpArgSerIlePheAspIleAspAsnIleAspIleGlyProGluArgArg 216
Db 1036 CGAAACTTCTGGCGCTCGTGGTTCGATGCTGACAGGGGTGTCTCTTGAGCCAGCTCGTCGA 1095
Qy 217 AlaTyrLeuGluGlnLysLeuGluArgValTrp-----GluGlnAsnThrGln 232
Db 1096 GATTATCTCAGTACAGTTCGGGACTGCTCACCCTCAACGCCCGCCCAACACACTTGT 1155
Qy 233 GlySerIleProTyrValProTyrLeuArgAlaPro-----TyrTyrVal--- 247
Db 1156 GGAGCAATTCCTTACTACATCTGTTGGG-ATGCTCAGAAGGCCCTTTTCTACGTCGAG 1214
Qy 248 -----TrpIleGlyArgLeuPro----- 253
Db 1215 TCTACCACAGAAATTGTGGAACACATCGATGCTATTGTGCGACAGAGTCCCCAAGATCATC 1274
Qy 253 ----- 253
Db 1275 ACCCGCTCTAGTGGGCAACATCTTCTCAGCAACAATGCGGTAGCCCTTTCAGTCGG 1334
Qy 254 -----SerValGlyHisAlaLeu-----HisGluGluArgValGlu 265
Db 1335 CCGCACAGCCCGGAGATGCGCTCAAAAGGCTTCAATACGCGCATCGAGAACAATCTCGCG 1394
Qy 266 ArgPro-ProMet----- 269
Db 1395 CAACCTTACCTCTCCCGTCTTCTTCTACCAAAATCACCACTGGAGGATCTACTAGACGA 1454
```

```
QY 270 -----PheProThrPheLeuTyrThrGlnSerTrpGluAs 282
Db 1455 TCAACTCCCGAAGCACACCCAGTTCAATGACGAGTACATCTACGCCCTTTACTCTGGGAAGA 1514
QY 282 pProGluProAspMetGluValMetGluIleAsnProLysAspThrValLeuThrLeuTh 302
Db 1515 CTGCGCGGTGACAGAGAATCTTAACTTCGGGCGGACGAGCTGCTCTACCATCAC 1574
QY 302 rSerGlyGlyCysAsnAlaLeuAsnLeuValGlnGlyAlaGlyGlnValValSerVa 322
Db 1575 CAGCGCGGCGACAACTTTCTTCTACCTGATGACAGAGTCCCGCTCGCGTGACCCCAT 1634
QY 322 lAspCysAsnProAlaGlnSerAlaLeuLeuGluLeuLysLysValAlaIleGlnLe 342
Db 1635 CGACCTAAACCCAGCGCAAAACACCTGCTTGAACCTCAAGTCCGCTCTTTACGACTCT 1694
QY 342 uGluPheGluAspValTrpGlnLeuPheGlyGluGlyValHisProArgIleGluLe 362
Db 1695 GGATTACCCGACGCTCTGGAAGATCTTCGGTGAGGGCAACACCCCGACTTTTCGCTCACT 1754
QY 362 uTyrGluLysLysLeuAlaProPheLeuSerGlnThrSerHisAsnPheTrpSerLysar 382
Db 1755 GCTCATCTCCAAACTCTCCCTCACCTTCGGCGCGGCTTCAATACCTGCTGATCCAA 1814
QY 382 gLeuTrpTyrPheGln-----HisGlyLeuTyrTyrGlnGlyGlyMetGlyLy 398
Db 1815 TGGGCACATATTTACCGACCTCGCGGCGCGCTCTATGATACCGGCGCTCCCGATA 1874
QY 398 sLeuCysTrpValLeuGlnCysLeuAlaValValLeuGlyLeuGlyLysThrValLysar 418
Db 1875 CGCTATCCGGTTCTTCGGTGGATTTCCACACTCTTTCTCGCGCTCCGCGCTCGCTCG 1934
QY 418 gLeuAlaAsnAlaProThrMetGluGluGlnArgArgLeuTrpAspSerAsnMetLeu1 438
Db 1935 ACTTCTCTACTCTCCACCTCGAAGGGCAAGCTTCACTACACCAAGAT----- 1989
QY 438 eHisPheValLysAsnGlyProLysProLeuValTrpLeuPheValLysPheValSer-- 457
Db 1990 -----CGTCCCTGT-----CTGCTCAACCGCTTCGTCAACGG 2021
QY 458 -LeuValLeuPheAsnLysAlaValLeuTrpPheGlyGlyGlyValProGlyLysGlnTy 477
Db 2022 CTTGGTCTCAGCTCCGACGCTTCTCTCTGTCGGCTTTTGGGCGTGCCCAAGAATCAAGT 2081
QY 477 xAlaLeuIleLysAlaasp----- 483
Db 2082 GGCTATGATCGAAGCCGACTACACCGCGCTTCTATCTCTCTCCACCAACCCCGAGCAG 2141
QY 484 -----GlyIleProIleGluAsnTyrIleAlaArgThrMetAspG1 497
Db 2142 CAAAGAAAACCAAGCGCGCCGAGCAATCTCCACTACACAACCTCCACCTTGATCC 2201
QY 497 yValAlaGluAsnSerHisValArgLysGlnAsnTyrPheTyrTyrAsnCysLeuThrG1 517
Db 2202 CGTTCTCTCCACCTCCACCTTCGCTCGGCAACCCCTTACTACTCTGCTGTGTCGTCGG 2261
QY 517 yLysPheLeuArgAspAsnCysProThrTyrLeuArgGluAlaAlaPheAlaThrLeuLy 537
Db 2262 ACAATACACACCGCAGTGCCATCCCGATTACCTTTCCCTGCGCCGCCACTCTATATCTAG 2321
QY 537 sSer---GlyValValAspAsnLeuThrValSerThrAsnPhePheMetGluGluLeu-- 555
Db 2322 CGTCTCTGGAGCCTTTGACCGCTTACGATCCACCGATGAATACAGGAGGTGTGGC 2381
QY 556 -----LysAlaArgThrTyrThrLysValIleLeuMetAspHisValAspTrpLeuAs 573
Db 2382 TAGGTTTCAGCGGGTACTTTGACAGTAGCGGTGGTGTGATGATGATGATGATGATGATG 2441
QY 573 pMetPro-----ValAlaAsnGluLeuAlaGluCy 593
Db 2442 TCGCCTCTCGCTGAGGAGAAAGGAAGGAGGCGGCAAGGCGAGGAGCAAGTGAGGAG 2501
```

```
QY 583 sLeuAlaLysGlnValAlaProGlyGlyIleValIleTrpArgSerAlaSerLeuSerPr 603
Db 2502 GTTGAATCCGGCGTTCGAAGTGGTGGAAGGTGTTGTTGAGAGCGCGGAGTGGAGCC 2561
QY 603 oProTyrAlaGluLeuIleGlnLysAlaGlyPheAspValArgCysIle----- 619
Db 2562 GTGGTATGTGAGGGTTCGAGGAGGTTTGGAGCAAGGTTTGGAGCAAGGAGGTGGTGTG 2621
QY 620 ----ArgArgAlaThrGlnGlyTyrMetAspArgValAsnMetTyrSerSerPheTyrMe 638
Db 2622 ATCCGGAAGGCGGACAGGAGTGTATTCACAGGTTGATATGATGCTAGTGTGAT 2681
QY 638 tAlaArgArg 641
Db 2682 CTGGGAAAG 2691
RESULT 5
US-10-620-914-48
; Sequence 48, Application US/10620914
; Publication No. US20040093639A1
; GENERAL INFORMATION:
; APPLICANT: Benning, Christoph
; APPLICANT: Riekhof, Wayne
; APPLICANT: Klug, Rouven
; TITLE OF INVENTION: Compositions and Methods for the Production of Betaine Lipids
; FILE REFERENCE: MSU-07769
; CURRENT APPLICATION NUMBER: US/10/620,914
; CURRENT FILING DATE: 2003-07-16
; PRIOR APPLICATION NUMBER: 10/118,495
; PRIOR FILING DATE: 2002-04-08
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 48
; LENGTH: 3427
; TYPE: DNA
; ORGANISM: Neurospora crassa
US-10-620-914-48
Alignment Scores:
Pred. No.: 5,11e-88 Length: 3427
Score: 804.00 Matches: 238
Percent Similarity: 43.2% Conservative: 112
Best Local Similarity: 29.4% Mismatches: 256
Query Match: 23.2% Indels: 205
DB: 7 Gaps: 24
US-10-620-914-45 (1-648) x US-10-620-914-48 (1-3427)
QY 19 LeuGluLysLeuLysLeuSerSerMetLysAspAspLeuThrValLeuArgHisMet--- 37
Db 776 TTGGAGGAGCTGGAGATG---GGCTTGAAGAAAGACAAAGGAATGGGAAGACATGTGTAA 832
QY 38 ---TrpPheGlySerLysLysGlyAspAspHisAlaAlaArgLeuGluGluSerPheTyrGly 56
Db 833 CGGTGCCCGGACAGGACCGGACCAAGGA-----GGAACCCATATGCG 877
QY 57 ProGlnAlaAlaPheAlaAlaArgLeuAla-----GluArgSerAsn 71
Db 878 TAGATGTACGCTCTCTCATCAGAACCTTGCAATGTGAAATCCCAACACGCTGACC 937
QY 72 LeuIleTrpValAspLeuGlyGlyThrGlyGluAsnValAspMetMetAlaAspTyr 91
Db 938 ATCGCTACAAAACAGGTGCGTGGGGGCACAGGCTGGAATATATCGAAGCCATGGCAAGTTT 997
QY 92 IleAspLeuAlaLys---PheLysSerIleTyrValValAspLeuCysHisSerLeuCys 110
Db 998 GTCACGCTCTGAAATCTTCAAGACTTTCAGTGTACCTAGTGAGCTTTTCGCCGCTCATTGT 1057
QY 111 GluValAlaLysLysLysAlaLysAlaLysGlyTrpLysAsnValGlnValValGluAla 130
Db 1058 GAAGTGGCTAGGAGGCGGTTTCCAGGCTGGGGTGGAGAGATGTGAGAGTTATCTGCAGC 1117
QY 131 AspAlaCysGlnPheAla-----ProProGluGlyThr----- 141
```



```

Db 669 GACCTCGCTAAGGCGCTCGCTG-----TTTTTTT-----TTCCGCTCGG 701
Qy 471 yValProGlyLysGlnTyr-----AlaLeuLeuLysAlaAspGlyLeuProleGluAs 489
Db 702 CATTCGCGCGCGAGTACGATCCCTGATCACCTCAGCGCGAGCCACCATGGCCAGCGT 761
Qy 489 nTyrIleAlaArg-----ThrMetAspGlyValAlaGluAsnSerHisValAr 505
Db 762 TCTGAAGCGCGCTGGAAGCTCGCTCGATTTTCCCTCGAATAAC-----810
Qy 505 gLysGlnAsnTyrPheTyrTyrAsnCysLeuThrGlyLysPheLeuArgAspAsn-----523
Db 811 -----AATTATTCGCTGGCAGGCTTTTGGCGCGCGCTATCCAAATCCCGTGAGGC 863
Qy 524 ----CysProThrTyrLeuArgGluAlaAlaPheAlaThrLeuLysSerGlyValValAs 542
Db 864 CGCCCTCGCGCGCTATCGAAGAGCAGAACTACGAAACCATCCGC---GGCAATATCGA 920
Qy 542 pAsnLeuThrValSerThrAsnPhePheMetGluGluLeu-----LysAlaArgTh 559
Db 921 CGCGTGGCATCCACCATGCCAATCTGATCGAATTCCTTCGCGGCAAGCAGCGGCAC 980
Qy 559 rTyrThrLysValIleLeuMetAspHisValAspTyrLeuAspMetProValAlaAsnG 579
Db 981 CGTCGATCGCTTCATCTGCTCGATGCGCAGGACTGGATGACCGATCACCAGCTCAACGC 1040
Qy 579 uLeuAlaGluCysLeuAlaLysGlnValAlaProGlyGlyIleValIleTrpArgSerAl 599
Db 1041 GCTGTGTCGGAATACAGCCGACCGCTCGCAGCGCGCGCTCATCTTCCGCAACGC 1100
Qy 599 a-----SerLeuSerProProTyrAlaGluLeuIleGlnLysAlaGlyPheAsp-V 616
Db 1101 CGCGAGCCGACGCTGCTGCGAGCGCGCTCGACCTCGCTGCTGACACGATGGGACTA 1160
Qy 616 alArgCysIleArgAlaThrGlnGlyTyrMetAspArgValAsnMetTyrSerSerP 636
Db 1161 TCAGGACGAGCGCTCGCGGAAATTCCTCGGCACGC-GACCGTTCGCGCATCTATGGCGCT 1219
Qy 636 heTyrMetAlaArgArgLysGlyAla 644
Db 1220 TCCACCTCTATGTGAAGCGCAGCGCA 1245

```

RESULT 10

```

US-10-118-495-28
; Sequence 28, Application US/10118495
; Publication No. US20030074688A1
; GENERAL INFORMATION:
; APPLICANT: Benning, Christoph
; APPLICANT: Riekhof, Wayne
; APPLICANT: Klug, Rouven
; TITLE OF INVENTION: Compositions and Methods for the Production of Betaine Lipids
; FILE REFERENCE: NSU-06897
; CURRENT APPLICATION NUMBER: US/10/118,495
; PRIOR FILING DATE: 2002-04-08
; PRIOR FILING DATE: 2001-04-13
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 28
; LENGTH: 1251
; TYPE: DNA
; ORGANISM: Agrobacterium tumefaciens
US-10-118-495-28

```

Alignment Scores:

Pred. No.:	2,86e-29	Length:	1251
Score:	326.50	Matches:	115
Percent Similarity:	47.6%	Conservative:	71
Best Local Similarity:	29.4%	Mismatches:	162
Query Match:	9.4%	Indels:	46
DB:	5	Gaps:	14

```

US-10-620-914-45 (1-648) x US-10-118-495-28 (1-1251)
Qy 275 LeuTyrThrGlnSerTyrGluAspProGluProAspMetGluValMetGluIleAsnPro 294
Db 130 GTCTATCCGAGATCTGGGAAGACCCGAGATCGATCGAAGCGATGGAGCTTGGCGAA 189
Qy 295 LysAspThrValLeuThrLeuThrSerGlyGlyCysAsnAlaLeuAsnLeuValGln 314
Db 190 GGCCACCGCATCTCACCATCGCTCCGCGCGTGAACATGCTGGCTCTCTCTCTCGCGC 249
Qy 315 GlyAlaGlyGlnValValSerValAspCysAsnProAlaGlnSerAlaLeuLeuGluLeu 334
Db 250 AACCCGCGCAGCATCGATGTGGACCTCAACCCGACACATCGCGCTGAACAAGATG 309
Qy 335 LysLysValAlaIleGlnLeuLeu---GluPheGluAspValTyrGlnLeuPheGlyGlu 353
Db 310 AACTCGCTGCTTCCGCCATCTGCGCGCCCATCAGATGTGGTGGCCACTTTCGGCGCG 369
Qy 354 GlyValHisProArgIleGluGluLeu-TyrGluLysLysLeuAlaProPheLeuSerG 373
Db 370 GCGCG-CACCCGCGCAGCAACAGCGTCGTTATGACCGTTTCATCGCCGAGCATCTGGATGC 428
Qy 373 nThrSerHisAsnPheTyrSerLys-----ArgLeuTyrTrp 386
Db 429 CACGACCAAGGCATCTACTGTGGAAGCGCACCCCTTTCGCGCGCGCTCGCATCTTCGGTGT 488
Qy 386 eGlnHisGlyLeuTyrTyrGlnGlyMetGlyLysLeuCysTyrTrpValLeuGlnCysLe 406
Db 489 CGACAGAACATCTACCGGACCGCTGCTCGCGCGCTTCATCGCGCGCGCCACATCAT 548
Qy 406 uAlaValValLeuGlyLeuGlyLysThrValLysArgLeuAlaAsnAlaProThrMetG 426
Db 549 GCGCGCGCTGCGCGTGAACCTCACC-----GAAATGGCAAGACCGCGCGCTGGA 602
Qy 426 uGluGlnArgArgLeuTyrAspSerAsnMetLeuIleHisPheValLysAsnGlyProLy 446
Db 603 CGAACAGCGCGCTTTTTCGACAGCAAGTTCGCGCGCTTTTC-----GACAA 650
Qy 446 sProLeuVal---TrpLeuPheValLysPheValSerLeuValLeuPheAsnLysAlaVa 465
Db 651 GCGGTGGTGGCTGGCTGCGAGCGCAAGCGCAAGAGCTCGCTT-----690
Qy 465 lLeuTyrPheGlyGlyValProGlyLysGlnTyr-----AlaLeuIleLysAl 482
Db 691 -----TTGCGCTTGGCATTTCGCGCGCCAGTATGACGAGCTGGCAAGCGCTTCCAG 743
Qy 482 aAspGlyIleProIleGluAsnTyrIleAlaArgThrMetAspGlyValAlaGluAsnSe 502
Db 744 CGACGCGC---ACGGTTGCTCTCGCTCAAGAGGCGCGCTGGAAGAGCTTCGCTGCAACTT 800
Qy 502 rHisValArgLysGlnAsnTyrPheTyrTyrAsnCysLeuThrGlyLysPheLeuArgAs 522
Db 801 CCGGCTC---AGGCACAATATTATTCGCTGCGAGCGCTTTCGCGCGCGCTTATCCGAGCC 857
Qy 522 pAsn-----CysProThrTyrLeuArgGluAlaAlaPheAlaThrLeuLysSerG 539
Db 858 GCATGAGGTGCGCTGCGCGCTTATCTCAAGCGCGAATATTACGAAAAGATCGCAACAA 917
Qy 539 yValValAspAsnLeuThrValSerThrAsnPhePheMetGluGluLeuLys-----556
Db 918 CACCGCG---CGGCTCGCGGTGCATCACGCGCACCTATACCGAGCTGCTTTTCCGCAAGCC 974
Qy 557 -AlaArgThrTyrThrLysValIleLeuMetAspHisValAspTyrLeuAspMetProVa 576
Db 975 GGCAAAATGGCGTGCACCGCTATATCTGCTCGATGCGCAGACTGGATGACGGATGTGCA 1034
Qy 576 lAlaAsnGluLeuAlaGluCysLeuLysGlnValAlaProGlyGlyIleValIleTr 596
Db 1035 GCTCAACAGTATGTCGCGAGATCAGCCGACATCGCGCATCGGGGCGACGCTCATCTT 1094
Qy 596 pArgSerAlaSerLeu-----SerProTyrTrpAlaGluLeuIleGlnLy 611
Db 1095 CCGCACGCGCGCGCA-AAAGAGCGTTCATCGAGGCGCGCTTTCGCGCGACATCGCAACC 1153

```

QY 611 sAlaGlyPheAspValArgCysLeuArgAlaThrGlnGlyTyrMetAspArgValas 631
Db 1154 AGTGGGTCTATCTCGAAGAGCGCTCCAAACA---ACTCAAC---GCCATGGACCGCTCGGC 1208
QY 631 nMetTyrSerPheTyrMetAlaArgArg 641
Db 1209 CATTTATGGCGGCTTCCATATCTACCAGAGG 1239
RESULT 11
US-10-620-914-28
; Sequence 28, Application US/10620914
; Publication No. US20040093639A1
; GENERAL INFORMATION:
; APPLICANT: Benning, Christoph
; APPLICANT: Riekhof, Wayne
; APPLICANT: Klug, Rouven
; TITLE OF INVENTION: Compositions and Methods for the Production of Betaine Lipids
; FILE REFERENCE: MSU-07769
; CURRENT APPLICATION NUMBER: US/10/620,914
; PRIOR FILING DATE: 2003-07-16
; PRIOR APPLICATION NUMBER: 10/118,495
; PRIOR FILING DATE: 2002-04-08
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 28
; LENGTH: 1251
; TYPE: DNA
; ORGANISM: Agrobacterium tumefaciens
US-10-620-914-28
Alignment Scores:
Pred. No.: 2,866-29 Length: 1251
Score: 326.50 Matches: 115
Percent Similarity: 47.6% Conservative: 71
Best Local Similarity: 29.4% Mismatches: 162
Query Match: 9.4% Indels: 46
DB: 7 Gaps: 14
US-10-620-914-45 (1-648) x US-10-620-914-28 (1-1251)
QY 275 LeuTyrThrGlnSerTrpGluAspProGluProAspMetGluValMetGluLeuLeuAsnPro 294
Db 130 GTCTATCCGAGATCTGGGAAGACCCCGAGATCGACATGAGCGATGGCGAA 189
QY 295 LysAspThrValLeuThrLeuThrSerGlyGlyCysAsnAlaLeuAsnLeuValGln 314
Db 190 GGCCACCGCATCGTACCATCGGCTCGCGGCTGCAACATGCTGGCTATCTCTCGGC 249
QY 315 GlyAlaGlyGlnValValSerValAspCysAsnProAlaGlnSerAlaLeuLeuLeu 334
Db 250 AACCCGGCCAGCATCGATGTGTGGACCTCAACCCGCCACCATCGGCTGAACAAGCTG 309
QY 335 LysLysValAlaLeuGlnGlnLeu---GluPheGluAspValTrpGlnLeuPheGlyGlu 353
Db 310 AAGCTCGCTCCCTCCCGCATCTCGCCGCCCATCAGGATGTGTGGCCACATTTCCGCGCG 369
QY 354 GlyValHisProGlnGluGluLeu-TyrGluLysLeuLeuAlaProPheLeuSerG1 373
Db 370 GCCGG-CACCCGGCAGCAACAGCTCGGTATGACCGGTTTCATCGCCGAGCATCTGGATGC 428
QY 373 nThrHisAsnPheTrpSerLys-----ArgLeuTrpTyrPh 386
Db 429 CACGACCAAGGCATCTGGTTCGAAGCCACCTTCGGCGCGGCTCGCATTTCCGTGT 488
QY 386 eGlnHisGlyLeuTyrTrpGlnGlyMetGlyLysLeuCysTyrValLeuGlnCysLe 406
Db 489 CGACAGGAACATCTACCGGACCGGCTCGCGGCTTCATCGCGCGGCGCCACATCAT 548
QY 406 uAlaValValLeuGlyLeuGlyLysThrValLysArgLeuAlaAsnAlaProThrMetG1 426
Db 549 GGCCCGCTCGACGGGCTGAACACTACC-----GAAATGGCCAAGACCGCGAGCTGGA 602

QY 426 uGluGlnArgArgLeuTrpAspSerAsnMetLeuLeuHisPheValLysAsnGlyProLy 446
Db 603 CGAACAGCCCGCTTTTGGACAGCAAGGTTCGCGCGCTTTTC-----GACAA 650
QY 446 sProLeuVal---TrpLeuPheValLysPheValSerLeuValLeuPheAsnLysAlaVa 465
Db 651 GCCGGTGGTGGCTGGCTGCTACGAGCGCAAGAGCTCGCTT----- 690
QY 465 lLeuTrpPheGlyGlyGlyValProGlyLysGlnTyr-----AlaLeuLeuLysAl 482
Db 691 -----TTCCGGCTTGGCATTCGCGCGCCAGTATGACGAGCTGGCAAGCTTTCCAG 743
QY 482 aAspGlyLeuProLeuGluAsnTyrIleAlaArgThrMetAspGlyValAlaGluAsnSe 502
Db 744 CGACGGC---ACGGTTGGCTCCCTCAAGGAGCGGCTGGAAAGACTTGCCTGCAACTT 800
QY 502 rHisValArgLysGlnAsnTyrPheTyrTrpAsnCysLeuThrGlyLysPheLeuArgAs 522
Db 801 CCGGCTC---AGGACACATTTTCCCTGGCAGGCGCTTGGCGCGGCTTATCCGAGCC 857
QY 522 pAsn-----CysProThrTyrLeuArgGluAlaAlaPheAlaThrLeuLysSerG1 539
Db 858 GCATGAGGTGGCTCGCGCTTATCTCAAGCGGAATATTACGAAAGATCCGCAACAA 917
QY 539 yValValAspAsnLeuThrValSerThrAsnPhePheMetGluGluLeuLys----- 556
Db 918 CACCGCG---CGCGTCGGGTGTCATCACGCCACCTATACCGAGCTGCTTTCCCGCAAGCC 974
QY 557 -AlaArgThrTyrThrLysValIleLeuMetAspHisValAspTrpLeuAspMetProVa 576
Db 975 GGCAATGGCTGCGACCGCTATCTGCTGATGCGAGGCTGGATGATCGGATGTGCA 1034
QY 576 lAlaAsnGluLeuAlaGluCysLeuAlaLysGlnValAlaProGlyGlyLeuValIleTr 596
Db 1035 GCTCAACAGATGTATGTCGCGATCAGCGCATCGCGCATCGGGCGCAGCGCTCATCTT 1094
QY 596 pArgSerAlaSerLeu-----SerProTyrAlaGluLeuLeuGlnLy 611
Db 1095 CGCACCGCGCGCGA-AAAGAGCGTTATCGAGGCGCGGCTTTCGCGCGACATCCGCAACC 1153
QY 611 sAlaGlyPheAspValArgCysIleArgAlaThrGlnGlyTyrMetAspArgValas 631
Db 1154 AGTGGGTCTATCTCGAAGAGCGCTCCAAACA---ACTCAAC---GCCATGGACCGCTCGGC 1208
QY 631 nMetTyrSerPheTyrMetAlaArgArg 641
Db 1209 CATTTATGGCGGCTTCCATATCTACCAGAGG 1239
RESULT 12
US-10-118-495-32
; Sequence 32, Application US/10118495
; Publication No. US20030074688A1
; GENERAL INFORMATION:
; APPLICANT: Benning, Christoph
; APPLICANT: Riekhof, Wayne
; APPLICANT: Klug, Rouven
; TITLE OF INVENTION: Compositions and Methods for the Production of Betaine Lipids
; FILE REFERENCE: MSU-06897
; CURRENT APPLICATION NUMBER: US/10/118,495
; CURRENT FILING DATE: 2002-04-08
; PRIOR APPLICATION NUMBER: 60/283,812
; PRIOR FILING DATE: 2001-04-13
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 32
; LENGTH: 1251
; TYPE: DNA
; ORGANISM: Sinorhizobium meliloti
US-10-118-495-32
Alignment Scores:
Pred. No.: 4,896-25 Length: 1251
Score: 292.50 Matches: 106

Db 976 CCGCGCGCTCAGTGCAGCGCTAGTGCTCTCTCGACGCACAGGACTGGATGACCGACCG 103
 Qy 576 ValAlaAsnGluLeuAlaGluCysLeuAlaLysGlnValAlaProGlyGlyLeuValIle 595
 Db 1036 CAGCTGAACGACCTCTGGACGGAGATCACCGCAGCCGCGCGCGGTGCTGATC 1095
 Qy 596 TrpArgSer-----AlaSerLeuSerProPro-----TyrAlaGluLeuIleGln 610
 Db 1096 TTCCGACACGGCGCGGAAGCAGCATCTCTCCGGGGCGCTCTCCACCACCCCTCTCGAT 1155
 Qy 611 LysAlaGlyPheAspValArgCysIleArgAlaGlnThrGlnGlyTyrMetAspArgVal 630
 Db 1156 CAGTGGTACTATGATGCCGAGACTTCGTGATGAGCTC-----GGCGCTGAAGACCGGTGC 1209
 Qy 631 AsnMetTyrSerSerPheTyrMetAlaArgArgLys 642
 Db 1210 GCGATCTATGGCGCTTCCACATCTACCGGAAGAA 1245
 RESULT 13
 US-10-620-914-32
 ; Sequence 32, Application US/10620914
 ; Publication No. US20040093639A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Benning, Christoph
 ; APPLICANT: Riekhof, Wayne
 ; APPLICANT: Klug, Rouven
 ; TITLE OF INVENTION: Compositions and Methods for the Production of Betaine L
 ; FILE REFERENCE: MSU-07769
 ; CURRENT APPLICATION NUMBER: US/10/620,914
 ; CURRENT FILING DATE: 2003-07-16
 ; PRIOR APPLICATION NUMBER: 10/118,495
 ; PRIOR FILING DATE: 2002-04-08
 ; NUMBER OF SEQ ID NOS: 52
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 32
 ; LENGTH: 1251
 ; TYPE: DNA
 ; ORGANISM: Sinorhizobium meliloti
 US-10-620-914-32
 Alignment Scores:
 Pred. No.: 4,89e-25 Length: 1251
 Score: 292.50
 Percent Similarity: 43.7%
 Best Local Similarity: 25.7%
 Query Match: 8.4%
 DB: 7
 US-10-620-914-45 (1-648) x US-10-620-914-32 (1-1251)
 Qy 263 ArgValGluArgProProMetPhePro-----ProThrPhe 274
 Db 74 AAGCTCTCTCCCGCGCGGTCTCTCCGAACGCCTGTTCGGGTGCTCTTTTCGGA-CTC 132
 Qy 275 LeuTyrThrGlnSerTrpGluAspProGluProAspMetGluValMetGluIleAsnPro 294
 Db 133 GTCTACCCGCAGATCTCGGAGGACCCGATTGTCACATGGAAGCGATGCAGATCCGTC 192
 Qy 295 LysAspThrValLeuThrLeuThrSerGlyGlyCysAsnAlaLeuAsnLeuValGln 314
 Db 193 GGACATCGGATCGTGAGCATCGGTTCGGCGGTGCAACATGTGACCTATCTCTCCGCC 252
 Qy 315 GlyAlaGlyGlnValSerValAspCysAsnProAlaGlnSerAlaLeuLeuGluLeu 334
 Db 253 GAGCCTGCCGGATAGACGTGGTTCGATCTCMACCCCATCATCGCGCTCAACCGGTG 312
 Qy 335 LysLysValAlaIleGlnGlnLeu---GluPheGluAspValTrpGlnLeuPheGly--- 352
 Db 313 AAGCTGTCTTCCTCCCACTCCCGAGCCACAAGGACGTGGTGGTTCCTCGCCGTC 372
 Qy 353 GluGlyValHisProArgIleGluGluLeuTyrGluLysLysLeuAlaProPheLeuSer 372
 Db 373 GAAGGTACGGGCACGAAT---GGCCAGGCGCTACGACGTGTTCCTCGCGCGAAGCTCGAT 429

Db 3047 AGCTTGAAGAGGATAAAATGGAGAGCAGTATCGATTCCATATCCGGGAGCCGGTTTA 3106
QY 246 -----TyrValTrrple-----GlyArgLeuProSerValGlyHisA 258
Db 3107 CGGCGCCTAAGTAACGAAGTTGGATGCAATGTGATGATGCGCTGTGGGCGCG 3166
QY 258 laLeuHisGluGluArgVal-----GluArgProProMetPhePro-ThrPhe 274
Db 3167 GCTGGTAAGAGAGCCTTATAGCGCATTTGAACCACTCCGGCTATCCGGGAGATATT 3226
QY 275 LeuTyrThrGlnSerTrpGluAspProGluProAspMetGluValMetGluLeuAsnPro 294
Db 3227 ATTTATCTGCGCAAAA----- 3241
QY 295 LysAspThrValLeuThrSerGlyGlyCysAsnAlaLeuAenLeuValGln 314
Db 3242 -----GGAATATCGATTATGTCTGCAATCACAGTC 3271
QY 315 GlyAlaGlyGlnValValSer-----ValAspCysAsnProAlaGlnSerAla 330
Db 3272 TCCCGTCGGAAGTGGTCAACGGGCATACGAGCTTATCTGCTCAACCACTATCCGCCAC 3331
QY 331 LeuLeuGluLeuLysLysValAlaIleGlnGlnLeuPheGluAspValTrpGlnLeu 350
Db 3332 ATTCGTGCTGTACGAAGAGTACACTGCTGCAATCGAC----- 3370
QY 351 PheGlyGluGlyValHisProArgIleGluLeuLeuTyrGluLysLysLeuAlaProPhe 370
Db 3371 -----ACATTATCCGGCACTGGCTGAA----- 3394
QY 371 LeuSerGlnThrSerHisAsnPheTrpSerLysArgLeu-----TrpTyrPheGlnHis 388
Db 3395 ATCTCATCAATGACGAAGAGTATGGCGGTAAACCCGACTGGACTGGCCATGAACAG 3454
QY 389 GlyLeuTyrTyrGlnGlyGlyMetGlyLysLeuCysTrpValLeuGln----- 404
Db 3455 GAT-----TTTCGCTGTGGT-----TGCTGGCTGACTGAGAAACCGAAACC 3496
QY 405 ---CysLeuAlaValValLeuGlyLysThrValLysArg----- 418
Db 3497 GCAATGAAGCCATTACCCGCAACTGACCGTAAATCTGGCGTACCTGATGCAACGC 3556
QY 419 -----LeuAlaAsnAlaProThrMetGluGluGlnArgArg----- 430
Db 3557 TCCGGGATGCTGTATTAATGATGCGCAGACCCGCTGATACATGATCCGATCACTGGAG 3616
QY 431 -----LeuTrpAspSerAsnMetLeuLeuHisPhe----- 440
Db 3617 TACGATAAATTTCCGGGAAATCAGTGAAGAGAATATTCTGAGTACCTTTGAGCAGCTACAC 3676
QY 441 -----ValLysAsnGlyProLysProLeuValTrp 450
Db 3677 CAGATAAAGGATGAAGTTTGGACGAGGATGATCAACCTCTTCAGAGGGTTGAGCTGG 3736
QY 451 LeuPhe-----ValLysPheValSerLeuValLeuPheAsnLysAlaVal 465
Db 3737 AATTACAAAACCAATTGCCCTGTAAATTTGGCAATAAATATATCGTCAACATCTGGTG 3796
QY 466 LeuTrpPheGlyGly-----ValProGlyLysGlnTyrAlaLeuLeuLysAla 482
Db 3797 AGGTGGGACCGGTGGGATGTCATCTTATCACCGGGCAACAGGAGAGTAGACTT---GCC 3853
QY 483 Asp-----GlyIleProIle-----GluAsn 489
Db 3854 GACCTGGAAGAAATGCTGATCTCTTTCAGCGGCAACCGATGCCCGCAACAGCGGAAC 3913
QY 490 TyrIleAlaArgThrMetAspGlyValAlaGluAsnSerHisValArgLysGlnAsnTyr 509
Db 3914 ATAATCTATTCATCTGGACGAT-----CACATCCGGTCT----- 3946
QY 510 PheTyrTyrAsnCysLeuThrGlyLysPheLeuArgAspAsn-----CysPro 525

Db 3947 -----CTTCAGGAAAGAAATGCTATGAAGATGAGATGTTTATTATAA 3991
QY 526 ThrTyrLeuArgGluAlaAlaPheAlaThrLeu---LysSerGlyValValAspAsnLeu 544
Db 3992 TACTTTAAGAGGGCTCTGGGCATCATCAGCTTCAGGAGCCAGAACTGGTTGACAGACTG 4051
QY 545 ThrValSerThrAsnPhePheMetGluGluLeuLysAlaArgThrTyrThrLysValIle 564
Db 4052 -----AATGATATTATTATGCAAGAACA----- 4072
QY 565 LeuMetAspHisValAspTrpLeuAspMetProValAlaAsnGluLeuAlaGluCys--- 583
Db 4073 CTATCTGACATCTGCGCAGTATGCTTTGCCAGTATGCTCAAGCTCGGAAATTTGTCG 4132
QY 584 -----LeuAlaLysGlnValAla-----ProGlyGlyIleVal 594
Db 4133 ATAAGATATGATCGTGGGCGCATATAACCGTTTCAGATGGCATACCCGCTGGCATACT- 4191
QY 595 IleTrpArg 597
Db 4192 GTCTGGAAG 4200

RESULT 15
US-10-684-141-57
; Sequence 57, Application US/10684141
; Publication No. US20050003536A1
; GENERAL INFORMATION:
; APPLICANT: Furusawa, Mitsuru
; TITLE OF INVENTION: METHOD AND SYSTEM FOR RAPIDLY CONFERRING
; FILE REFERENCE: 690116.401
; CURRENT APPLICATION NUMBER: US/10/684,141
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 57
; LENGTH: 7119
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-684-141-57

Alignment Scores:
Pred. No.: 0.0106 Length: 7119
Score: 123.00 Matches: 144
Percent Similarity: 33.7% Conservative: 83
Best Local Similarity: 21.4% Mismatches: 214
Query Match: 3.6% Indels: 234
DB: 8 Gaps: 36

US-10-620-914-45 (1-648) x US-10-684-141-57 (1-7119)

QY 107 HisSerLeuCysGluValAlaLysLysLysAlaLysAlaLys---GlyTrpLysAsnVal 125
Db 1307 CATATCTC-----AAGGAGCTGCCAGGCCAAACTTGGCTATGACCCCTGTA 1354
QY 126 GlnValValGluAlaAspAlaCysGlnPheAlaProGluGlyThrAlaThrLeuIle 145
Db 1355 GAGCTGGACCTGAGGACATGTCTGATGCCC---ACTGAACAGCCCGACAGCTCTGGCC 1411
QY 146 ThrPheSerTyrSer----- 150
Db 1412 ACTTACTCAGTGTACAGATGCTGTGGCTTACTTACTCCTGTACATGAAATACGTCCACCCC 1471
QY 150 ----- 150
Db 1472 TTCATATTCGCCCTGTGCAACCATATTCCCATGGAACCTGATGAGGTGTCGGGAAGGGC 1531
QY 151 -----LeuThrMetIleProProPheHis---AsnValIle--- 161
Db 1532 TCCGGGACACTGTGTGAAGCCTTGTGCTGATGCTGCAAGCTTTCCATGCGCAACATATCTTC 1591
QY 162 -----AspGlnAlaCysSerGlnAspGlyLeuValGlyValAla 177

Db 1592 CCCAATTAAGCAAGCAGGAGTTCAACAAGCTGACAGATGATGGCCACCGTGTAGATGCT 1651
Qy 178 AspPheTyrValSerGly-----LysTyrAspLeu 187
Db 1652 GAGACCTACGTTGGGGCCACGTCGGAGCCACTAGAGTCTGGTGTCTTCAGAAGTGATATC 1711
Qy 188 Pro-----LeuArgGlnMetProTyrSerArgPhePheTyrArgSerIlePheAsp 205
Db 1712 CCTCGCGGTTAGGATGAATCTCGCAGCCTTTGATTTCTGTGCTCAACGAGTCGAGAAG 1771
Qy 206 IleAspAsnIleAspIleGlyProGluArgArgAlaTyrLeuGluGln-----LysLeu 223
Db 1772 ACTATGCCACCGCATTTAGAAAGAAAGAGAGTGGCTGTGGAAACAGCCCAACTTT 1831
Qy 224 GluArgValTrpGluGln-----AsnThrGlnGlySerIleProTyrValPro---Trp 240
Db 1832 CAAGAGGTGTGAGCAGATTAAGACCAAGCTCACCTCCCTAAAGATGTTCTCTAACAGA 1891
Qy 241 LeuArgAlaPro-----TyrTyrValTrpIleGlyArgLeu---ProSerVal 255
Db 1892 ATTGAATGTCTCTTAATCTATCATCTAGATGTGGGGCCCATGTATCTCTAACATAATCTTT 1951
Qy 256 ---GlyHisAlaLeuHisGluGluArgValGlu-ArgProProMetPheProProThrPh 274
Db 1952 ACCAACCGCTACAGCCTTCTGCCATAGTAGAGGAGCCACCTGTGTGCTGTGACTTC 2011
Qy 274 eLeuTyrThrGlnSer-----TrpGluAspProGluProAspMetGluVa 289
Db 2012 AATAAGCCTGAGCAAGTGTCTCAGAGGAAGTGGCTGGC----- 2051
Qy 289 lMetGluIleAsnProLysAspThrValLeuThrLeuThrSerGlyGlyCysAsnAlaLe 309
Db 2052 -----AGTGGAGGGGAGAATTCA-- 2069
Qy 309 uAsnLeuLeuValGlnGlyAlaGlyGlnValValSerValAspCysAsnProAlaGlnSe 329
Db 2070 -----TGCCAGCCAT-CCGAG 2085
Qy 329 rAlaLeuLeuGluLeuLysLysValAlaIleGlnGlnLeuGluPheGluAspValTrpG1 349
Db 2086 TGAATACCATCGGATTCAG-----CATCAGTGGAGTCGGAGAAGTTTCCCCC 2133
Qy 349 nLeuPheGlyGlyGlyValHisProArgIleGluGluLeu----- 362
Db 2134 TTTGTTTCCAGAGGGGCCAGCAGCGGCTTTTCACGAGCTGTCCCGTGAAGAACAGGCTAA 2193
Qy 363 -TyrGluLys----LysLeuAlaProPhe----- 370
Db 2194 ATATGAGAAGAGGAGGCTGGCAGATTATTCGCGAAAGCCCTATAAGAGATCCATGTGAC 2253
Qy 371 -----LeuSerGlnThrSerHisAsnPheTrpSerLysAr 382
Db 2254 CAAGGTAGAGAACGCTTAACCTACCTCCAGCGGGAAGAACTCATTTATGTGGACAC 2313
Qy 382 gLeuTrpTyrPheGlnHisGlyLeuTyrTyrGlnGlyMetGlyLysLeuCysTrpVa 402
Db 2314 AGTGGCGGCTTCAGACAGCGGCTATGATTTCAAAGGACTGCACAAGGTGTGAAGAA 2373
Qy 402 lLeuGlnCysLeuAlaValValLeuGlyLeuGlyLysThrValLysArgLeuAlaAsnAl 422
Db 2374 GAAGCTCTCGGAGCTGTGAGGTGGGCGATGATCAGAGGTGAAGCGCTGCAAGAAC-- 2431
Qy 422 aProThrMetGluGluGlnArgLeuTrpAspSerAsnMetLeuIleHis----- 439
Db 2432 -----ATGAGATCTCTTACGATTCTACGAGTGGCTCACAAGTGAT 2475
Qy 440 -----PheValLysAsnGlyProLysProLeuValTrpLeuPh 452
Db 2476 CCTGAACCTCTTACGGCTATGTCATGCGCAAGGAGCTCGC-----TGGTATTC 2526
Qy 452 eValLysPheValSerLeuValLeuPheAsnLysAlaValLeuTrpPheGlyGlyVa 472
Db 2527 CATGGAGATGGTGGTATCGTCTGCTTACA-----GGACCAACAT 2568

Qy 472 lProGlyLysGlnTyrAlaLeuIleLysAlaAspGlyIleProIle----- 487
Db 2569 CATCCCAAGCAAGAGAACTGATTGACGAGATCGGAGCGCTTTAGAAATTGGACACGGA 2628
Qy 488 -----GluAsnTyrIleAlaArgThrMetAs 496
Db 2629 CGGAATATGTGTGCTCTACCCCAATAGCTTTCTCTGAAANATTTGTCTCAAGACA----- 2693
Qy 496 pGlyValAlaGluAsnSerHisValArgLysGlnAsnTyrPheTyrTyrAsnCysLeuTh 516
Db 2684 -----ACCAATCGGAGAAACCAAACTGACCATCTCTATCTCTGCTGCTGCTT 2733
Qy 516 rGlyLysPheLeuArgAspAsnCysProThrTyr-----LeuArgGluAlaAl 532
Db 2734 GAACATCATGTGTCAAGGAAGCTTTTACCAACCACTAGTACCAGGAACCTAACAGAGCCTTC 2793
Qy 532 aPheAlaThrLeuLysSerGlyValValAspAsnLeuThrValSerThrAsnPhePheMe 552
Db 2794 GTCTCTCACTAT-----GTCCACCCTCTGAGAAATAGTATCTT 2832
Qy 552 tGluGluLeuLysAlaArgThrTyrThrLysValIleLeuMetAspHisValAspTcPle 572
Db 2833 TTTTGAAGTCGATGA--CCATACCTTGTGTATGATCCTT----- 2869
Qy 572 uAspMetProValAlaAsnGluLeuAlaGluCysLeuAlaLysGlnValAlaProGlyG1 592
Db 2870 -----CCAGCCTCCAAGGAAGCAAGAGCTGAAGAAAGATATGCT----- 2914
Qy 592 ylleValIleTrpArgSerAlaSerLeuSerProProTyrAlaGluLeuIleGlnLysAl 612
Db 2915 ---GTGTTCAATGAAGATGTTCTCTG-----GCTGAACCTG-----AA 2949
Qy 612 aGlyPheAspValArgCysIleArgAlaThrGlnGlyTyrMetAspArgValAsnMe 632
Db 2950 AGTTTGTAGGTGAAA-----CCCCGAGGAGGATTCGACGCTGATTAAANAT 2994
Qy 632 tTyr---SerSerPheTyrMetAlaArgArgLysGlyAla 644
Db 2995 ATTCCAGTCTCTCAGTTTTTTGAGGCGCTTCTCTCAAGGCGCAGC 3034

Search completed: March 14, 2006, 07:07:47
Job time : 1419 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 14, 2006, 06:44:17 ; Search time 1226 Seconds
(without alignments)
1219.517 Million cell updates/sec

Title: US-10-620-914-45
Perfect score: 3463
Sequence: 1 MSGRDRPASYTKNFSLE.....RVNYSFYMARKGAKGN 648

Scoring table:

	BLOSUM62	Xgapop 10.0 , Xgapext 0.5	Ygapop 10.0 , Ygapext 0.5	Fgapop 6.0 , Fgapext 7.0	Delop 6.0 , Delext 7.0
Sequence:	1	1	1	1	1

Searched: 7673375 seqs, 1153648444 residues

Total number of hits satisfying chosen parameters: 15346750

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q=/abes/ABSWEB spool/US10620914/runat 13032006 102052 6964/app query.fasta.1
-DB=Published Applications NA New -QWMT=fastap -SUFFIX=p2n.rnpbn -MINMATCH=0.1
-LOOPCL=0 -LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-THR_MIN=0 -ALIGN=15 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-MAXLEN=2000000000 -HOST=abes05h
-USER=US10620914 @CCN 1.1 431 @runat 13032006 102052 6964 -NCPU=6 -ICPU=3
-NO MMAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA New:

1:	/cgn2_6/ptodata/2/pubpna/US08 NEW PUB.seq.*
2:	/cgn2_6/ptodata/2/pubpna/US06 NEW PUB.seq.*
3:	/cgn2_6/ptodata/2/pubpna/US07 NEW PUB.seq.*
4:	/cgn2_6/ptodata/2/pubpna/US09 NEW PUB.seq.*
5:	/cgn2_6/ptodata/2/pubpna/US09 NEW PUB.seq.*
6:	/cgn2_6/ptodata/2/pubpna/US10 NEW PUB.seq.*
7:	/cgn2_6/ptodata/2/pubpna/US10 NEW PUB.seq.*
8:	/cgn2_6/ptodata/2/pubpna/US10 NEW PUB.seq.*
9:	/cgn2_6/ptodata/2/pubpna/US11 NEW PUB.seq.*
10:	/cgn2_6/ptodata/2/pubpna/US11 NEW PUB.seq.*
11:	/cgn2_6/ptodata/2/pubpna/US11 NEW PUB.seq.*
12:	/cgn2_6/ptodata/2/pubpna/US11 NEW PUB.seq.*
13:	/cgn2_6/ptodata/2/pubpna/US60 NEW PUB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	118.5	3.4	1155	12 US-11-234-786-373	Sequence 373, App
2	118.5	3.4	2000	12 US-11-234-786-374	Sequence 374, App
3	118.5	3.4	2040	12 US-11-234-786-375	Sequence 375, App
4	115.5	3.3	1512	12 US-11-234-786-368	Sequence 368, App

ALIGNMENTS

RESULT 1

US-11-234-786-373
; Sequence 373, Application US/11234786
; Publication No. US20060024301A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yugu
; APPLICANT: Reed, Steven G.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.
; TITLE OF INVENTION: PROSTATE-SPECIFIC POLYPEPTIDES AND FUSION
; FILE OF INVENTION: POLYPEPTIDES THEREOF
; FILE REFERENCE: 210121.427C31
; CURRENT APPLICATION NUMBER: US/11/234,786
; CURRENT FILING DATE: 2005-09-23
; PRIOR APPLICATION NUMBER: US 09/568,857
; PRIOR FILING DATE: 2000-05-09

```
; PRIOR APPLICATION NUMBER: US 09/536,857
; PRIOR FILING DATE: 2000-05-27
; PRIOR APPLICATION NUMBER: US 09/483,672
; PRIOR FILING DATE: 2000-01-14
; PRIOR APPLICATION NUMBER: US 09/439,313
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: US 09/352,616
; PRIOR FILING DATE: 1999-07-13
; PRIOR APPLICATION NUMBER: US 09/288,946
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: US 09/232,149
; PRIOR FILING DATE: 1999-01-15
; PRIOR APPLICATION NUMBER: US 09/159,812
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: US 09/115,453
; PRIOR FILING DATE: 1998-07-14
; PRIOR APPLICATION NUMBER: US 09/030,607
; PRIOR FILING DATE: 1998-02-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 701
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 373
; LENGTH: 1155
; TYPE: DNA
; ORGANISM: Homo sapien
US-11-234-786-373

Alignment Scores:
Pred. No.: 0.0507 Length: 1155
Score: 118.50 Matches: 66
Percent Similarity: 37.4% Conservative: 19
Best Local Similarity: 29.1% Mismatches: 81
Query Match: 3.4% Indels: 61
DB: 12 Gaps: 11

US-10-620-914-45 (1-648) x US-11-234-786-373 (1-1155)

QY 187 LeuProLeuArgGlnMetProTTP-----SerArgArgPhePheTrpArgSerIlePhe 204
Db 204 CTTCCCTGCTGCGGGGGAGTGGCAAGACAGCAAGTGGGGCGCTTCGGAGA-----254

QY 205 AspIleAspAsnIleAspIleGlyProGluArgAlaTyrLeuGluGlnLysLeuGlu 224
Db 255 -----CCACGACGACTCTGCTAT-----GAA 275

QY 225 ArgValTrpGluGlnAsnThrGlnGlySerIleProTyrValProTrpLeuArgAlaPro 244
Db 276 GACACTCAGGAACAAGATGGGCAAGTGGTCTGCCACTGCTTCCCTGCTGCGAGGGGAG 335

QY 245 TyrTyrValTrpIleGlyArgLeuProSerValGly---HisAlaLeuHisGluGluArg 263
Db 336 CGCAAGACGCAAGTGGGGCCCTTGGGAGACTGAGTGCAGTGCCTTCATGAGGCCAG 395

QY 264 ValGluArgProMetPheProProThrPheLeuTyrThrGlnSerTrp-GluAsp-- 282
Db 396 GTACCACGTCGG-----TGAGAGAGATCT 419

QY 283 -----ProGluProAspMetGluValMe 290
Db 420 GGCAAGCTCCACAGAGCTGCTGGTGGGTAAAGTCCCGAAGAGGATCTCATGCGCAT 479

QY 290 t-----GluIleAsnProLysAsp-----ThrValLeuThrLeuTh 302
Db 480 GCTCAGGGACACTGACGTGACGAGCAAGAGGACAGCAAGAAAGAGGAGCTCTCTACATCTGGC 539

QY 302 rSer-----GlyGlyCysAsnAlaLeuAsnLeuLeuValGlnGlyAlaGlyGlnValVa 320
Db 540 CTCTGCCAATGGGAATTCAGAGAGTAGTAAAACTCTCTGCTGACAGACGATGTCAACTTAA 599

QY 320 lSerValAspCysAsnProAlaGlnSerAlaLeuLeuLeuLysLysValIleGln 340
Db 600 TGTCTTGTGAC---AACAAAGAGGAGGACGACTCTGATAAAG-----GCCGTACA 644
```

```
QY 340 nGlnLeuGluPheGluAspValTrpGlnLeuPheGlyGluGlyValHisProArgIleG1 360
Db 645 ATGCCAGGAAGATGATGCGTTAATGTTCTGGAAACATGGCACTGATCCAAATATTC 704

QY 360 uGluLeuTyr-GluLysLysLeuAlaProPheLeuSerGlnThrSerHisAsnPheTrpS 380
Db 705 AGATGAGTATGAAATACCACTCTGCACCTACGCTATCTATATGAAGATAAATTAATGGC 764

QY 380 erIysArgLeuTrpTyr 385
Db 765 CAAAGCACTGCTCTTAT 781
```

RESULT 2

```
US-11-234-786-374
; Sequence 374, Application US/11234786
; Publication No. US20060024301A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Reed, Steven G.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.
; TITLE OF INVENTION: PROSTATE-SPECIFIC POLYPEPTIDES AND FUSION
; TITLE OF INVENTION: POLYPEPTIDES THEREOF
; FILE REFERENCE: 210121.427C31
; CURRENT APPLICATION NUMBER: US/11/234,786
; PRIOR FILING DATE: 2005-09-23
; PRIOR APPLICATION NUMBER: US 09/568,857
; PRIOR FILING DATE: 2000-05-09
; PRIOR APPLICATION NUMBER: US 09/536,857
; PRIOR FILING DATE: 2000-05-27
; PRIOR APPLICATION NUMBER: US 09/483,672
; PRIOR FILING DATE: 2000-01-14
; PRIOR APPLICATION NUMBER: US 09/439,313
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: US 09/352,616
; PRIOR FILING DATE: 1999-07-13
; PRIOR APPLICATION NUMBER: US 09/288,946
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: US 09/232,149
; PRIOR FILING DATE: 1999-01-15
; PRIOR APPLICATION NUMBER: US 09/159,812
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: US 09/115,453
; PRIOR FILING DATE: 1998-07-14
; PRIOR APPLICATION NUMBER: US 09/030,607
; PRIOR FILING DATE: 1998-02-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 701
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 374
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Homo sapien
US-11-234-786-374

Alignment Scores:
Pred. No.: 0.105 Length: 2000
Score: 118.50 Matches: 66
Percent Similarity: 37.4% Conservative: 19
Best Local Similarity: 29.1% Mismatches: 81
Query Match: 3.4% Indels: 61
```

DB: 12 Gaps: 11

US-10-620-914-45 (1-648) x US-11-234-786-374 (1-2000)

187 LeuProLeuArgGlnMetProTyr-----SerArgArgPhePheTrpArgSerIlePhe 204
|||||
204 CTTCCCTGCTGCAGGGGAGTGGCAAGACAGCGTGGCGCTTCTGGAGA-----254
205 AspIleAspAsnIleAspIleGlyProGluArgAlaTyrLeuGluGlnLysLeuGlu 224
|||||
255 -----CCACGACGACTCTGCTAT-----GAA 275
225 ArgValTrpGluGlnAsnThrGlnGlySerIleProTyrValProTrpLeuArgAlaPro 244
|||||
276 GACACTCAGGAACAAGATGGCAAGTGTCTCCCTGCTGCGAGGGGAG 335
245 TyrTyrValTrpIleGlyArgLeuProSerValGly---HisAlaLeuHisGluGluArg 263
|||||
336 CGGCAAGACGAGGTGGCGCTTGGGGAGCTACGATGACGATGCTTCATGGAGCCAG 395
264 ValGluArgProProMetPheProThrPheLeuTyrThrGlnSerTrp-GluAsp-- 282
|||||
396 GTACCACGTCG-----TGGAGAAGATCT 419
283 -----ProGluProAspMetGluValMe 290
420 GGACAAGCTCCACAGAGCTGCTGGTGGGGTAAAGTCCCGACAGAAAGATCTCATCGTCAT 479
290 t-----GluIleAsnProLysAsp-----ThrValLeuThrLeuTh 302
480 GCTCAGGGACACTGACGTGACCAAGAGGACCAAGAGGACGACTGCTCTACATCTGGC 539
302 rSer-----GlyGlyCysAsnAlaLeuAsnLeuValGlnGlnGlyAlaGlyGlnValVa 320
540 CTCTGCCAATGGGAATTCAGAGTAGTATAAAGTCCCTGCTGGACAGACGATGTCAACTTAA 599
320 lSerValAspCysAsnProAlaGlnSerAlaLeuLeuGluLysValAlaIleG1 340
|||||
600 TGTCTTGGC---AACAAAAGAGGACGCTCTGATAAAG-----GCCGTACA 644
340 nGlnLeuGluPheGluAspValTrpGlnLeuPheGlyGluGlyValHisProArgIleG1 360
|||||
645 ATGCCAGGAGATGAATGTGCGTTAATGTGCTGGACATGGCAGTCAATATTC 704
360 uGluLeuTyr-GluLysLysLeuAlaProPheLeuSerGlnThrSerHisAsnPheTrpS 380
|||||
705 AGATGAGTATGGAATACCACTCTGCACGTACGCTATCTATAAGTAAATTAATGGC 764
380 erLysArgLeuTrpTyr 385
|||||
765 CAAAGCACTGCTCTTAT 781

RESULT 3

US-11-234-786-375
; Sequence 375, Application US/11234786
; Publication No. US20060024301A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Reed, Steven G.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darriack
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.

TITLE OF INVENTION: PROSTATE-SPECIFIC POLYPEPTIDES AND FUSION
FILE REFERENCE: POLYPEPTIDES THEREOF
CURRENT APPLICATION NUMBER: US/11/234,786
CURRENT FILING DATE: 2005-09-23
PRIOR APPLICATION NUMBER: US 09/568,857
PRIOR FILING DATE: 2000-05-09
PRIOR APPLICATION NUMBER: US 09/536,857
PRIOR FILING DATE: 2000-05-27
PRIOR APPLICATION NUMBER: US 09/483,672
PRIOR FILING DATE: 2000-01-14
PRIOR APPLICATION NUMBER: US 09/439,313
PRIOR FILING DATE: 1999-11-12
PRIOR APPLICATION NUMBER: US 09/352,616
PRIOR FILING DATE: 1999-07-13
PRIOR APPLICATION NUMBER: US 09/288,946
PRIOR FILING DATE: 1999-04-09
PRIOR APPLICATION NUMBER: US 09/232,149
PRIOR FILING DATE: 1999-01-15
PRIOR APPLICATION NUMBER: US 09/159,812
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: US 09/115,453
PRIOR FILING DATE: 1998-07-14
PRIOR APPLICATION NUMBER: US 09/030,607
PRIOR FILING DATE: 1998-02-25
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 701
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 375
LENGTH: 2040
TYPE: DNA
ORGANISM: Homo sapien
US-11-234-786-375

Alignment Scores:
Pred. No.: 0.108 Length: 2040
Score: 118.50 Matches: 66
Percent Similarity: 37.4% Conservative: 19
Best Local Similarity: 29.1% Mismatches: 81
Query Match: 3.4% Indels: 61
DB: 12 Gaps: 11

US-10-620-914-45 (1-648) x US-11-234-786-375 (1-2040)

QY 187 LeuProLeuArgGlnMetProTyr-----SerArgArgPhePheTrpArgSerIlePhe 204
DB 204 CTTCCCTGCTGCAGGGGAGTGGCAAGACAGCGTGGCGCTTCTGGAGA-----254
QY 205 AspIleAspAsnIleAspIleGlyProGluArgAlaTyrLeuGluGlnLysLeuGlu 224
DB 255 -----CCACGACGACTCTGCTAT-----GAA 275
QY 225 ArgValTrpGluGlnAsnThrGlnGlySerIleProTyrValProTrpLeuArgAlaPro 244
DB 276 GACACTCAGGAACAAGATGGCAAGTGTGCTGCCACTGCTCCCTGCTGCGAGGGGAG 335
QY 245 TyrTyrValTrpIleGlyArgLeuProSerValGly---HisAlaLeuHisGluGluArg 263
DB 336 CGGCAAGACGAGGTGGCGCTTGGGGAGCTACGATGACGATGCTTCATGGAGCCAG 395
QY 264 ValGluArgProProMetPheProThrPheLeuTyrThrGlnSerTrp-GluAsp-- 282
DB 396 GTACCACGTCG-----TGGAGAAGATCT 419
QY 283 -----ProGluProAspMetGluValMe 290
DB 420 GGACAAGCTCCACAGAGCTGCTGGTGGGGTAAAGTCCCGACAGAAAGATCTCATCGTCAT 479
QY 290 t-----GluIleAsnProLysAsp-----ThrValLeuThrLeuTh 302
DB 480 GCTCAGGGACACTGACGTGACCAAGAGGACCAAGAGGACGACTGCTCTACATCTGGC 539
QY 302 rSer-----GlyGlyCysAsnAlaLeuAsnLeuValGlnGlnGlyAlaGlyGlnValVa 320

```
Db 540 CTCTGCCAATGGGAATTCAGAACTAGTAAACTCTCTGGACAGACGATGTCAACTTAA 599
QY 320 lSerValAspCysAsnProAlaGlnSerAlaLeuLeuGluLeuLysValAlaIleG1 340
Db 600 TGTCTCTTGAC--AACAAAAGAGGACAGCTCTGTATAAAG-----GCCGTACA 644
QY 340 nGlnLeuGluPheGluAspValTrpGlnLeuPheGlyGluGlyValHisProArgIleG1 360
Db 645 ATGCCAGGAAGATGAATGTCGGTTAATGTTCTGGAACTGGCACTGATCCAATATATCC 704
QY 360 uGluLeuTyr-GluLysLysLeuAlaProPheLeuSerGlnThrSerHisAsnPheTrpS 380
Db 705 AGATGAGTATGGAATACCACTCTGCACCTACGCTATCTATAATGAAGATAAATTAATGC 764
QY 380 erLysArgLeuTrpTyr 385
Db 765 CAAAGCACTGCTCTTAT 781
```

RESULT 4

```
US-11-234-786-368
; Sequence 368, Application US/11234786
; Publication No. US20060024301A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Reed, Steven G.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.
; TITLE OF INVENTION: PROSTATE-SPECIFIC POLYPEPTIDES AND FUSION
; FILE REFERENCE: 210121.427C31
; CURRENT APPLICATION NUMBER: US/11/234,786
; PRIOR FILING DATE: 2005-09-23
; PRIOR APPLICATION NUMBER: US 09/568,857
; PRIOR FILING DATE: 2000-05-09
; PRIOR APPLICATION NUMBER: US 09/536,857
; PRIOR FILING DATE: 2000-05-27
; PRIOR APPLICATION NUMBER: US 09/483,672
; PRIOR FILING DATE: 2000-01-14
; PRIOR APPLICATION NUMBER: US 09/439,313
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: US 09/352,616
; PRIOR FILING DATE: 1999-07-13
; PRIOR APPLICATION NUMBER: US 09/288,946
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: US 09/232,149
; PRIOR FILING DATE: 1999-01-15
; PRIOR APPLICATION NUMBER: US 09/159,812
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: US 09/115,453
; PRIOR FILING DATE: 1998-07-14
; PRIOR APPLICATION NUMBER: US 09/030,607
; PRIOR FILING DATE: 1998-02-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 701
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 368
; LENGTH: 1512
; TYPE: DNA
; ORGANISM: Homo sapien
US-11-234-786-368
```

```
Alignment Scores:
Pred. No.: 0.149 Length: 1512
Score: 115.50 Matches: 65
Percent Similarity: 36.6% Conservative: 18
Best Local Similarity: 28.6% Mismatches: 83
Query Match: 3.3% Indels: 61
DB: 12 Gaps: 11

US-10-620-914-45 (1-648) x US-11-234-786-368 (1-1512)
QY 187 LeuProLeuArgGlnMetProTrp-----SerArgArgPhePheTrpArgSerIlePhe 204
Db 426 CTTCCCTCTGTCAGGGGAGTGGCAAGACGACGACGTCGCGCTCTCGGAGA----- 476
QY 205 AspIleAspAsnIleAspIleGlyProGluArgArgAlaTyrLeuGluGlnLysLeuGlu 224
Db 477 -----CCACGACGAYTCTGCTAT-----GAA 497
QY 225 ArgValTrpGluGlnAsnThrGlnGlySerIleProTyrValProTrpLeuArgAlaPro 244
Db 498 GACACTCAGGACACAGATGGCAAGTGGTGTGCTGCCACTGCTTCCCTGCTGCAGGGGAG 557
QY 245 TyrTyrValTrpIleGlyArgLeuProSerValGly---HisAlaLeuHisGluGluArg 263
Db 558 CRGCAAGAGCAAGGTGGCGCTTGGGGAGACTACGATGACAGTGCCTTCATGGAGCCAG 617
QY 264 ValGluArgProProMetPheProThrPheLeuTyrThrGlnSerTrp-GluAsp-- 282
Db 618 GTACCACGTCCTG-----TGGAGAAGATCT 641
QY 283 -----ProGluProAspMetGluValMe 290
Db 642 GGCAAGCTCCACAGAGCTGCCTGGTGGGTAAAGTCCCAAGAAAGATCTCATGTCAT 701
QY 290 t-----GluIleAsnProLysAsp-----ThrValLeuThrLeuTh 302
Db 702 GCTCAGGGACACTGACGTGTAACAAGAGGACAAAGAGGAGGACTGCTCTACATCTGGC 761
QY 302 rSer-----GlyGlyCysAsnAlaLeuLeuValGlnGlyAlaGlyGlnValVa 320
Db 762 CTCTGCCAATGGGAATTCAGAACTAGTAAAACTCTGTGGACAGACGATGTCAACTAA 821
QY 320 lSerValAspCysAsnProAlaGlnSerAlaLeuLeuLysLysValAlaIleG1 340
Db 822 TGTCTCTTGAC--AACAAAAGAGGACAGCTCTGAYAAAG-----GCCGTACA 866
QY 340 nGlnLeuGluPheGluAspValTrpGlnLeuPheGlyGlyGlyValHisProArgIleG1 360
Db 867 ATGCCAGGAAGATGAATGTCGGTTAATGTTCTGGAACTGGCACTGATCCAATATATCC 926
QY 360 uGluLeuTyr-GluLysLysLeuAlaProPheLeuSerGlnThrSerHisAsnPheTrpS 380
Db 927 AGATGAGTATGGAATACCACTCTRCACCTAATGCTCTCTCTCTCTCTCTCTCTCTCTCT 986
QY 380 erLysArgLeuTrpTyr 385
Db 987 CAAAGCACTGCTCTTAT 1003

RESULT 5
US-11-234-786-369
; Sequence 369, Application US/11234786
; Publication No. US20060024301A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Reed, Steven G.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
```

; APPLICANT: Stolk, John A.
 ; APPLICANT: Day, Craig H.
 ; APPLICANT: Vedvick, Thomas S.
 ; APPLICANT: Carter, Darrick
 ; APPLICANT: Li, Samuel X.
 ; APPLICANT: Wang, Aijun
 ; APPLICANT: Skeiky, Yasir A.
 ; TITLE OF INVENTION: PROSTATE-SPECIFIC POLYPEPTIDES AND FUSION
 ; FILE REFERENCE: 210121.427C31
 ; CURRENT APPLICATION NUMBER: US/11/234,786
 ; PRIOR FILING DATE: 2005-09-23
 ; PRIOR APPLICATION NUMBER: US 09/568,857
 ; PRIOR FILING DATE: 2000-05-09
 ; PRIOR APPLICATION NUMBER: US 09/536,857
 ; PRIOR FILING DATE: 2000-05-27
 ; PRIOR APPLICATION NUMBER: US 09/483,672
 ; PRIOR FILING DATE: 2000-01-14
 ; PRIOR APPLICATION NUMBER: US 09/439,313
 ; PRIOR FILING DATE: 1999-11-12
 ; PRIOR APPLICATION NUMBER: US 09/352,616
 ; PRIOR FILING DATE: 1999-07-13
 ; PRIOR APPLICATION NUMBER: US 09/288,946
 ; PRIOR FILING DATE: 1999-04-09
 ; PRIOR APPLICATION NUMBER: US 09/232,149
 ; PRIOR FILING DATE: 1999-01-15
 ; PRIOR APPLICATION NUMBER: US 09/159,812
 ; PRIOR FILING DATE: 1998-09-23
 ; PRIOR APPLICATION NUMBER: US 09/115,453
 ; PRIOR FILING DATE: 1998-07-14
 ; PRIOR APPLICATION NUMBER: US 09/030,607
 ; PRIOR FILING DATE: 1998-02-25
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 701
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 369
 ; LENGTH: 1853
 ; TYPE: DNA
 ; ORGANISM: Homo sapien
 US-11-234-786-369

Alignment Scores:
 Pred. No.: 0.83 Length: 1853
 Score: 109.50 Matches: 64
 Percent Similarity: 36.1% Conservative: 18
 Best Local Similarity: 28.2% Mismatches: 84
 Query Match: 3.2% Indels: 61
 DB: 12 Gaps: 11

US-10-620-914-45 (1-648) x US-11-234-786-369 (1-1853)
 QY 187 LeuProLeuArgGlnMetProTyr-----SerArgArgPhePheTyrArgSerIlePhe 204
 DB 426 CTTCCTCTCTGCGGGGAGTGGCAAGACGACGCTGGGGCTTCTGGAGA----- 476
 QY 205 AspIleAspAsnIleAspIleGlyProGluArgArgAlaTyrLeuGluGlnLysLeuGlu 224
 DB 477 -----CCACGACGATCTGCTAT-----GAA 497
 QY 225 ArgValTyrGluGlnAsnThrGlnGlySerIleProTyrValProTyrLeuArgAlaPro 244
 DB 498 GACACTCAGGAACAAGATGGCAAGTGGCTGCCACTCTCCCTGCTGCGAGGGGGAG 557
 QY 245 TyrTyrValTyrIleGlyArgLeuProSerValGly---HisAlaLeuHisGluGluArg 263
 DB 558 CRGCAAGACGACGCTGGCGCTGGGGAGACTGACGATGACAGGCGCTTCATGAKKCCCG 617
 QY 264 ValGluArgProProMetPheProProThrPheLeuTyrThrGlnSerTyr-GluAsp-- 282
 DB 618 GTACCACGCTCC-----TGAGAGAGATCT 641
 QY 283 -----ProGluProAspMetGluValMe 290
 |||||
 |||||

Db 642 GGACAAGCTCCACAGAGCTGCCTGGTGGGGTAAAGTCCCGAGAAAGATCTCATCGTCAT 701
 QY 290 t-----GluIleAsnProLysAsp-----ThrValLeuThrLeuTh 302
 Db 702 GCTCAGGGACACKAGVGTGACCAAGARGGACCAAGCAAGAGGACTGCTCTACATCTGGC 761
 QY 302 rSer-----GlyGlyCysAsnAlaLeuAsnLeuValGlnGlyAlaGlyGlnValva 320
 Db 762 CTCTGCCAATGGGAATTCAGAAAGTAGTAAACTCTGTGGACAGACGATGTCACACTAA 821
 QY 320 lSerValAspCysAsnProAlaGlnSerAlaLeuLeuGluLeuLysValAlaIleG1 340
 Db 822 TGTCTTGTGAC---AACAAAAAGAGGACGCTCTGAYAAAG-----GCCGTACA 866
 QY 340 nGlnLeuGluPheGluAspValTyrGlnLeuPheGlyGluGlyValHisProArgIleG1 360
 Db 867 ATGCCAGGAAGATGAATGTCGTTAATGTTGCTGGAAACATGGCACTGATCCAAATATTC 926
 QY 360 uGluLeuTyr-GluLysLysLeuAlaProPheLeuSerGlnThrSerHisAsnPheTyrPs 380
 Db 927 AGATGAGTATGGAAATACCACCTCTRCACCTAYGCTRTCTAYATGAAGATAAATTAATGGC 986
 QY 380 erIysArgLeuTyrTyr 385
 Db 987 CAAGCACTGCTCTTAT 1003

RESULT 6
 US-11-234-786-531
 ; Sequence 531, Application US/11234786
 ; Publication No. US20060024301A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Xu, Jiangchun
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Mitcham, Jennifer L.
 ; APPLICANT: Harlocker, Susan L.
 ; APPLICANT: Jiang, Yugu
 ; APPLICANT: Reed, Steven G.
 ; APPLICANT: Kalos, Michael D.
 ; APPLICANT: Fanger, Gary R.
 ; APPLICANT: Retter, Marc W.
 ; APPLICANT: Stolk, John A.
 ; APPLICANT: Day, Craig H.
 ; APPLICANT: Vedvick, Thomas S.
 ; APPLICANT: Carter, Darrick
 ; APPLICANT: Li, Samuel X.
 ; APPLICANT: Wang, Aijun
 ; APPLICANT: Skeiky, Yasir A.
 ; TITLE OF INVENTION: PROSTATE-SPECIFIC POLYPEPTIDES AND FUSION
 ; FILE REFERENCE: 210121.427C31
 ; CURRENT APPLICATION NUMBER: US/11/234,786
 ; CURRENT FILING DATE: 2005-09-23
 ; PRIOR APPLICATION NUMBER: US 09/568,857
 ; PRIOR FILING DATE: 2000-05-09
 ; PRIOR APPLICATION NUMBER: US 09/536,857
 ; PRIOR FILING DATE: 2000-05-27
 ; PRIOR APPLICATION NUMBER: US 09/483,672
 ; PRIOR FILING DATE: 2000-01-14
 ; PRIOR APPLICATION NUMBER: US 09/439,313
 ; PRIOR FILING DATE: 1999-11-12
 ; PRIOR APPLICATION NUMBER: US 09/352,616
 ; PRIOR FILING DATE: 1999-07-13
 ; PRIOR APPLICATION NUMBER: US 09/288,946
 ; PRIOR FILING DATE: 1999-04-09
 ; PRIOR APPLICATION NUMBER: US 09/232,149
 ; PRIOR FILING DATE: 1999-01-15
 ; PRIOR APPLICATION NUMBER: US 09/159,812
 ; PRIOR FILING DATE: 1998-09-23
 ; PRIOR APPLICATION NUMBER: US 09/115,453
 ; PRIOR FILING DATE: 1998-07-14
 ; PRIOR APPLICATION NUMBER: US 09/030,607
 ; PRIOR FILING DATE: 1998-02-25
 ; Remaining Prior Application data removed - See File Wrapper or PALM.

QY 283 -----ProGluProAspMetGluValMet-----GluLeuAsnProLysAsp-- 296
Db 1008 AGTCCCGAGAAAGATCTCATGCTCATGCTCAGGACACCGATGTGAACAAGAGGACAA 1067
QY 297 -----ThrValLeuThrLeuThrSer-----GlyGlyCysAsnAlaLeuAsnLe 311
Db 1068 GCAAAAGAGAGACTCTCATCTGGCCTCTGCCAATGGGAATTCAGAAAGTAGTAAACT 1127
QY 311 uLeuValGlnGlyAlaGlyGlnValValSerValAspCysAsnProAlaGlnSerAlaLe 331
Db 1128 CGTGTGGACAGACGATGTCACCTTAATGTCCTTGAC---AACAARAAGAGACAGCTCT 1184
QY 331 uLeuGluLeuLysValAlaLeuGlnGlnLeuGluPheGluAspValTrpGlnLeuPh 351
Db 1185 GACAAG-----GCCGTACCAATGCCAGGAAGATGAATGTGCGTAAATGTTGCT 1232
QY 351 eGlyGluGlyValHisProGlyGluGluLeuValTrpGlyGlyLeuAlaProPheL 371
Db 1233 GGAACATGGCACTGATCCCAATATTCAGATGATGGAATACCACTCTACACTATGC 1292
QY 371 euSerGlnThrSerHisAsnPheTrpSerLysArgLeuTrpTyr 385
Db 1293 TGCTCAATGAAGATAAATTAATGGCAAGCACTCTCTAT 1336

RESULT 8

US-10-932-182A-4215
; Sequence 4215, Application US/10932182A
; Publication No. US20060046253A1
; GENERAL INFORMATION:
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: 030685-043
; CURRENT APPLICATION NUMBER: US/10/932, 182A
; CURRENT FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 4215
; LENGTH: 3006
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-10-932-182A-4215

Alignment Scores:
Pred. No.: 5.92 Length: 3006
Score: 104.00 Matches: 57
Percent Similarity: 39.8% Conservative: 43
Best Local Similarity: 22.7% Mismatches: 79
Query Match: 3.0% Indels: 72
DB: 7 Gaps: 15

US-10-620-914-45 (1-648) x US-10-932-182A-4215 (1-3006)

QY 407 AlaValValLeuGlyLeuGlyLysThrValLysArgLeuAlaAsnAlaProThrMetGlu 426
Db 1672 TCAATTTAAGAGACTGAATATGATATCCCTCAATTTGGTCAATGTT----- 1719
QY 427 GluGlnArgLeuTrpAspSerAsnMetLeuLeuHisPheValLysAsnGlyProLys 446
Db 1720 -----AGTCTCTAGTTAGTGTATTT----- 1740
QY 447 ProLeuValTrpLeuPheValLysPheValSerLeuValLeu-----PheAsn 462
Db 1741 -----TGGATCTTTTTCGAATTTTAGGAGTACAGATTTTTCGAAGTTTCATTCGA 1791
QY 463 LysAlaValLeuTrpPheGlyGlyValProGlyLys-----GlnTrpAlaLeuLeu 480
Db 1792 AGCAATGTGTGGTATTAATCCGAGAGACCTTACTGATACATATCATACAGCCTG---- 1848

QY 481 LysAlaAspGlyLeuProLeuGluAsnTyrLeuAlaArgThrMetAspGlyValAlaGlu 500
Db 1849 -----CAGTTTGGCGCGGCCCATCTTGACCCAGTGACAAAA 1884
QY 501 AsnSerHisValArgLysGlnAsnTyrPheTyrAsnCysLeuThrGlyLys----- 518
Db 1885 -----CAAGACAAAAACTATATCTATGAGGATGGTACTGAAGGTCTGTTTCA 1932
QY 519 -----PheLeuArgAspAsnCysProThrTyrLeuArgGluAlaAlaPheAlaThrLeu 536
Db 1933 AAGGCTTCTT-----TGCCCAACATATTCGAAGTGTCTTCTAATGCTAATCCG 1983
QY 537 LysSerGlyValVal-----AspAsnLeuThrValSerThrAsnPheMetGluGlu 554
Db 1984 TATAACGGTAGAATAGCTTTTGATAATATTTGTCATTCGATGGAACCTTGCTTTGTGTCATA 2043
QY 555 LeuLysAlaArgThrTyrThrLysValLe-----LeuMetAspHisValAspTrpLeu 572
Db 2044 ATGAGTGTAAACACCTTCACTGATTTGATGATATATACGATGGAT----- 2088
QY 573 AspMetProValAlaAsnGluLeuAlaGluCysLeuAlaLysGlnValAlaProGlyGly 592
Db 2089 -----TCAGATGAATGGCTGCATGTTGTTTTTATCGTTTGTATTTTGTGTC 2136
QY 593 IleValIleTrpArgSerAlaSerLeuSerProProTyrAlaGluLeuLeuGlnLysAla 612
Db 2137 TTAACCTATTGG-----CTACTGAATTTACTTATTGTCAGTCTCTGGTG-----TCT 2181
QY 613 GlyPheAspValArgCysIleArgAlaThrGlnGlyTyrMetAspArgValAsnMet 632
Db 2182 TCTTTCGAATA-----GCTAACGGAGGTACAGAAGAGAGAACAAATA 2226
QY 633 TyrSerSer-----PheTyrMetAlaArg 640
Db 2227 TATGGTTCAAAAAACATGGGTTATATGCGCGT 2259

RESULT 9

US-10-932-182A-4215
; Sequence 4215, Application US/10932182A
; Publication No. US20060046253A1
; GENERAL INFORMATION:
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: 030685-043
; CURRENT APPLICATION NUMBER: US/10/932, 182A
; CURRENT FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 4215
; LENGTH: 3006
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-10-932-182A-4215

Alignment Scores:
Pred. No.: 5.92 Length: 3006
Score: 104.00 Matches: 57
Percent Similarity: 39.8% Conservative: 43
Best Local Similarity: 22.7% Mismatches: 79
Query Match: 3.0% Indels: 72
DB: 7 Gaps: 15

US-10-620-914-45 (1-648) x US-10-932-182A-4215 (1-3006)

QY 407 AlaValValLeuGlyLeuGlyLysThrValLysArgLeuAlaAsnAlaProThrMetGlu 426
Db 1672 TCAATTTAAGAGACTGAATATGATATCCCTCAATTTGGTCAATGTT----- 1719
QY 427 GluGlnArgLeuTrpAspSerAsnMetLeuLeuHisPheValLysAsnGlyProLys 446

```

Db      1720  -----AGTCTCTAGTTAGTGTATTT-----1740
QY      447  ProLeuValTrpLeuPheValSerLeuValLeu-----PheAsn 462
Db      1741  -----TGGATCTTTTGGAAATTTTAGGAGTACAGATTTTCAAGGTTCAATTCGA 1791
QY      463  LysAlaValLeuTrpPheGlyGlyValProGlyLys-----GlnTyrAlaLeuLeu 480
Db      1792  AGCAATGTGTGGTATTAATTCGGAAGACCCTACTGATACATATCAATACACCGTG---1848
QY      481  LysAlaAspGlyIleProIleGluAsnTyrIleAlaArgThrMetAspGlyLysAlaGlu 500
Db      1849  -----CAGTTTGGCGGGCCATCTTGACCAGTCAGCAAAA 1884
QY      501  AsnSerHisValArgLysGlnAsnTyrPheTyrTyrAsnCysLeuThrGlyLys-----518
Db      1885  -----CAAAAGACAAAACATATATCTATGAGGTGGTACTGAAGGTCCTGTTTCA 1932
QY      519  -----PheLeuArgAspAsnCysProThrTyrLeuArgGluAlaAlaPheAlaThrLeu 536
Db      1933  AAGGGCTTTCTT-----TGCCACAAATATCCAAAGTGTCTTCTAATGCTAATCCG 1983
QY      537  LysSerGlyValVal-----AspAsnLeuThrValSerThrAsnPhePheMetGluGlu 554
Db      1984  TATAACGGTAGAATTAGCTTTGATATATATGTCAAATTCGATGGAACCTTGCTTTGTCA 2043
QY      555  LeuLysAlaArgThrTyrThrLysValIle-----LeuMetAspHisValAspTrpLeu 572
Db      2044  ATGAGTGTCAACACCTTCACTGATTTGATGATATTATATACGATGGAT-----2088
QY      573  AspMetProValAlaAsnGluLeuAlaGluCysLeuAlaLysGlnValAlaProGlyGly 592
Db      2089  -----TCAGATGAATGGCTGCAGTTGTTTATGTTTATGTTGTTATTTTGTGTC 2136
QY      593  IleValIleTrpArgSerAlaSerLeuSerProProTyrAlaGluLeuIleGlnLysAla 612
Db      2137  TTAACATTTGG-----CTACGAATTTACTTATTCAGTCTCTGGTG-----TCT 2181
QY      613  GlyPheAspValArgCysIleArgAlaThrGlnGlyTyrMetAspArgValAsnMet 632
Db      2182  TCTTTTCAATA-----GCTAACGGAGGTACAAAGAGAGAAACAATA 2226
QY      633  TyrSerSer-----PheTyrMetAlaArg 640
Db      2227  TATGGTTCAAAAAACATGGGTTTATTTCCCGGT 2259

```

RESULT 10

```

US-10-932-182A-2428
; Sequence 2428, Application US/10932182A
; Publication No. US20060046253A1
; GENERAL INFORMATION:
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIOHKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: 030685-043
; CURRENT APPLICATION NUMBER: US/10/932,182A
; CURRENT FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 2428
; LENGTH: 1857
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-10-932-182A-2428

Alignment Scores:
Pred. No.: 5.08 Length: 1857
Score: 102.00 Matches: 68
Percent Similarity: 33.1% Conservative: 36

```

```

Best Local Similarity: 21.7% Mismatches: 116
Query Match: 2.9% Indels: 94
Db: 7 Gaps: 15

US-10-620-914-45 (1-648) x US-10-932-182A-2428 (1-1857)

QY      84  AsnValAspMetMetAlaAspTyrIleAspLeuAlaLysPheLysSerIleTyrValVal 103
Db      604  AATGTGAGATGTTCCGTGACTTACCTGAAGTAAAAAACCAATGAGAAGATAT-----657
QY      104  AspLeuCysHisSerLeuCysGluValAlaLysLysLysAlaLysLysGlyTyrLys 123
Db      658  -----CATTTCTGAAAAACACAAATTCCAAATAATGGTAAATGGCTTCATGAAA 708
QY      124  AsnValGln-----ValValGluAlaAspAlaCysGlnPheAlaProPro 138
Db      709  GTGGTTCAAGGTTCTCTGTGGATGGCTTTAAGCGCACTAGGGATGAATACTTCCCTCCT 768
QY      139  Glu-----GlyThrAlaThrLeuIleThrPheSerTyrSer 150
Db      769  TCATATGTTTTGGATAGAAAGACTTTCGGCGCAAGGCTCTTTATTATTTCAGGTCACATTAC 828
QY      151  LeuThrMetIleProProPheHisAsn-----ValIleAspGln 163
Db      829  TTATTCTTACTAGGTTTATTACAGATTTAAGTATTATGCTGCTCGCTGGACCAATTCGAA 888
QY      164  AlaCysSerTyrLeuSerGlnAspGlyLeuValGlyValAlaAspPheTyrValSerGly 183
Db      889  GCATCTGTGTTATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 924
QY      184  LysTyrAspLeuProLeuArgGlnMetProTyrSerArgPhePheTyrArgSerIle 203
Db      925  ---TATGATCCAAAGACTCAAAAGATCAGATGGGATCGT-----960
QY      204  PheAspIleAspAsnIleAspIleGlyProGluArgArgAlaTyrLeuGluGlnLysLeu 223
Db      961  -----GTTGAAATATTCACATTTTCGACCGTAGAACCGCAACAAAGTACACGTGAATG 1014
QY      224  GluArgValTrpGluGlnAsnThrGlnGlySerIleProTyrValProTyrLeuArgAla 243
Db      1015  TTGGAGCGCATGGATATGAACTAAC-----AAGTGGCTGAAGTAC 1056
QY      244  ProTyrTyrValTrpIleGlyArgLeuProSerValGlyHisAlaLeuHisGluGluArg 263
Db      1057  TCGGTTTATTTG-----CGTGTACTAAAAAGGCT-----GAGAAGCCC 1095
QY      264  ValGluArgProMetPheProThrPheLeuTyrThrGlnSerTrpGluAspPro 283
Db      1096  GGTTCCTCGTTCAACTTTGTTTC-----ACTTCTTAACTTCAGCATTTTGGCAGGTACC 1149
QY      284  GluProAspMetGluValMetGluIleAsnProLysAspThrValLeuThrLeuThrSer 303
Db      1150  AGACCTGGTTAC-----TACCTGACGTTTGGCCACG 1179
QY      304  GlyGlyCysAsnAlaLeuAsnLeuValGlnGlyAlaGlyGlnValSerValAsp 323
Db      1180  GGGGCG-----TTGTACCAACATGTCGTAAATTTACAGCGTAAT 1221
QY      324  CysAsnProAlaGlnSerAlaLeuLeuLysLysValAlaIleGlnGlnLeuGlu 343
Db      1222  TTCAGGCCA-----ATGTTCTTCAGGAAGATGTTGTTAGCTCTACCC 1266
QY      344  PheGluAspValTrpGlnLeuPheGlyValHisProArgIleGluGluLeuTyr 363
Db      1267  TCCAAATAATTTATGATCTCATCGGC-----ATATAC 1299
QY      364  GluLysLysLeuAla-----ProPheLeu 371
Db      1300  GCGATCAAGTTAGCAATTTGGCTATATGTTACACCAATTCATC 1341

```

RESULT 11

```

US-10-932-182A-2428
; Sequence 2428, Application US/10932182A

```

```

; Publication No. US20060046253A1
; GENERAL INFORMATION:
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHIRO
; APPLICANT: NAKAMURA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIOHKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: 030685-043
; CURRENT APPLICATION NUMBER: US/10/932,182A
; CURRENT FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 2428
; LENGTH: 1857
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-10-932-182A-2428

Alignment Scores:
Pred. No.: 5.08 Length: 1857
Score: 102.00 Matches: 68
Percent Similarity: 33.1% Conservative: 36
Best Local Similarity: 21.7% Mismatches: 116
Query Match: 2.9% Indels: 94
DB: 7 Gaps: 15

US-10-620-914-45 (1-648) x US-10-932-182A-2428 (1-1857)

QY 84 AsnValAspMetMetAlaAspTyrIleAspLeuAlaLysPheLysSerIleTyrValVal 103
DB 604 AATTGGAGATGTTCCGTGACTTACCTGAAAGTAAACCAATGAGAAGATAT----- 657
QY 104 AspLeuCyHisSerLeuCyGluValAlaLysLysLysLysLysLysGlyTyrLys 123
DB 658 -----CATTCGGAAAAGACAAATTCAAAATGTAATGGTAAATGGCTTCATGAAA 708
QY 124 AsnValGln-----ValValGluAlaAspAlaCysGlnPheAlaProPro 138
DB 709 GTGGTTCAAGGTCCTCTGTGGATGGCTTTAAGCGCACTAGGGATGAATACATTCCTCCT 768
QY 139 Glu-----GlyThrAlaThrLeuIleThrPheSerTyrSer 150
DB 769 TCATATGTTTGGATGAAAGACTTTCGGGCAAGGTCCTTTATTTTCAGGTACATAC 828
QY 151 LeuThrMetIleProProPheHisAsn-----ValIleAspGln 163
DB 829 TTATTTCTAGTGTTTTATTCACAGATTAAGTATATGCTGCTGGACCATTTCCGAA 888
QY 164 AlaCysSerTyrLeuSerGlnAspGlyLeuValGlyValAlaAspPheTyrValSerGly 183
DB 889 GCATCTGTGTATTATGTGATTAGGTATAATGGT----- 924
QY 184 LysTyrAspLeuProLeuArgGlnMetProTyrSerArgArgPheThrArgSerIle 203
DB 925 ---TATGATCAAGACTCAAAAGATCAGATGGGATCGT----- 960
QY 204 PheAspIleAspAsnIleAspIleGlyProGluArgAlaTyrLeuGluGlnLysLeu 223
DB 961 -----GTAGAAATATGATATTGGACCGGTAGAACGCAAAAGTACACGTGAATG 1014
QY 224 GluArgValTrpGluGlnAnthrGlnGlySerIleProTyrValProTyrLeuArgAla 243
DB 1015 TTGGAGCATGGAATATGAACATAAC-----AAGTGGCTGAAGTAC 1056
QY 244 ProTyrTyrValTrpIleGlyArgLeuProSerValGlyHisAlaLeuHisGluGluArg 263
DB 1057 TCCGTTTATTG-----CGTGTACTAAAAAGGGT-----GAGAAGCCC 1095
QY 264 ValGluArgProMetPheProThrPheLeuTyrThrGlnSerTrpGluAspPro 283
DB 1096 GGTTCCTCAACTTTGTTTC-----ACTTCTTAACCTTACGATTTTGGACCGGTACC 1149

```

```

QY 284 GluProAspMetGluValMetGluIleAsnProLysAspThrValLeuThrLeuThrSer 303
DB 1150 AGACCTGGTTAC-----TACCTGACGTTTGGCAGC 1179
QY 304 GlyGlyCyAsnAlaLeuAsnLeuValGlnGlyAlaGlyGlnValValSerValAsp 323
DB 1180 GGGCG-----TTGTACCAAAACATGTGTTAAATTTACAGGCGTAAT 1221
QY 324 CysAsnProAlaGlnSerAlaLeuLeuGluLeuLysLysValAlaIleGlnGlnLeuGlu 343
DB 1222 TTCAGGCCA-----ATGTTCTTACGGGAAGATGGTGTAGCCCTCTACCC 1266
QY 344 PheGluAspValTrpGlnLeuPheGlyGlyGlyValHisProArgIleGluGluLeuTyr 363
DB 1267 TCCAAAATAATTTATGATCTCATCGGC-----ATATAC 1299
QY 364 GluLysLysLeuAla-----ProPheLeu 371
DB 1300 GCGATCAAGTTAGCATTTGGCTATATGTTACCAACCATTCATC 1341

RESULT 12
US-11-124-367A-21
; Sequence 21, Application US/11124367A
; Publication No. US20060024700A1
; GENERAL INFORMATION:
; APPLICANT: Michele Cargill
; APPLICANT: Hongjin Huang
; TITLE OF INVENTION: Genetic Polymorphisms Associated with
; FILE REFERENCE: CL001519.ORD
; CURRENT APPLICATION NUMBER: US/11/124,367A
; CURRENT FILING DATE: 2005-05-09
; PRIOR APPLICATION NUMBER: US 60/568,846
; PRIOR FILING DATE: 2004-05-07
; PRIOR APPLICATION NUMBER: US 60/582,609
; PRIOR FILING DATE: 2004-06-25
; PRIOR APPLICATION NUMBER: US 60/599,554
; PRIOR FILING DATE: 2004-08-09
; NUMBER OF SEQ ID NOS: 34460
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 3042
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-11-124-367A-21

Alignment Scores:
Pred. No.: 9.75 Length: 3042
Score: 102.00 Matches: 90
Percent Similarity: 33.3% Conservative: 54
Best Local Similarity: 20.8% Mismatches: 163
Query Match: 2.9% Indels: 126
DB: 12 Gaps: 19

US-10-620-914-45 (1-648) x US-11-124-367A-21 (1-3042)

QY 12 TyrThrLysLysAsnPheSerLeuGluLysLeuSerSerMetLysAspLeu 31
DB 199 TTTTACCGGACTGAGTTTCAGATCGTGAATTCAAA---GCCACATGTGCAACTACTG 255
QY 32 ThrValLeuArgHisMetTrpPheGlySerLysLysGlyAspAspHisAlaAla----- 49
DB 256 GCATATCTAAAGCACCTC-----AAAGGGCAAAACGAGCGCCCTGGAA 300
QY 50 -----ArgLeuGluSerPheTyrGlyProGlnAlaAla-----AlaPheAlaAla 64
DB 301 TCGTTACGTAAGCTGAAGATGTAATCCAGCAAGAGCATGCTGACCGAGCAAGAAATCAGA 360
QY 65 ArgLeuAlaGluArgSerAsnLeuIleTrpValAspLeuGlyGlyThrGlyGluAsn 84
DB 361 AGTCTGGTCACCTGGGAAACTATGCTGGGTCTACTATCACAATGGCCGCACTCTCAGAC 420
QY 85 ValAspMetMetAlaAspTyrIleAsp-----LeuAlaLysPheLysSerIleTyrVal 102

```

```
Db 421 GTTCAGATTTATGTAGACAGGTGAACATGTCTGTGAGAAGTTTTCAGTCCCTATAGA 480
Qy 103 ValAspLeuCysHisSerLeuCysGluValAlaLysLysLysLysLysLysGlyTrp 122
Db 481 ATTTGAGAGTCCAGAGCTTGACTGTGAGGAAGGGTGGACACGGTTAAAGTGTGGAGAAAC 540
Qy 123 LysAsn-----ValGlnValGluAlaAspAlaCysGlnPheAlaPro-----Pro 138
Db 541 CAAATGAACAGCAAGGTGTCTTTGAGAAGCTCTGAAAGAGCCCAAGAACCCCA 600
Qy 139 GluGlyThrAlaThrLeuLeuThrPheSerTyrSerLeuThrMetLeuProPheHis 158
Db 601 GAATTCACCTCTGGAGTGGCAATAGCAAGCTACCGTCTGCACCACTGGCCACCATCTCAG 660
Qy 159 AsnValLeuAspGlnAlaCysSerTyrLeuSerGlnAspGlyLeuValGlyValAlaAsp 178
Db 661 AAGCCATTGAC----- 672
Qy 179 PheTyrValSerGlyLysTyrAspLeuProLeuArgGlnMetProTyrSerArgArgPhe 198
Db 673 -----CCTCTGAGGCAAGCC----- 687
Qy 199 PheTrpArgSerIlePheAspIleAspAsnIleAspIleGlyProGluArg----- 215
Db 688 -----ATTCGGCTGAATCCTGCACACCAAGTACCTT 717
Qy 216 ArgAlaTyrLeuGluGlnLysLeuGluArgValTrpGluGlnAsnThrGlnGlySerIle 235
Db 718 AAAAGTCTCTGGCTCTGAAGCTTCTAAGATCGTGAAGAA----- 759
Qy 236 ProTyrValProTyrLeuArgAlaProTyrTyrValTrpIleGlyArgLeuProSerVal 255
Db 760 -----GCTGAGAGGAGGTGAA 777
Qy 256 GlyHisAlaLeuHisGluGluArgValGluArgProPro-----MetPhePro 271
Db 778 GGAGAGAAGTTAGTTGAAGAAGCCTTGAGAAAGCCCGCAGGTGTAAACAGATGTTCTTCGC 837
Qy 272 ProThrPheLeuTyrThrGlnSerTrpGluAspProGluProAspMetGluVal----- 289
Db 838 AGTGCAGCCAAAGTTTATATCAAGAAAGATGAGCCAGCAAAAGCGATTGAAGCTTTAAA 897
Qy 290 -----MetGluIleAsnProLysAspThrValLeuThrLeuThrSerGlyGlyCysAsn 307
Db 898 AAGCTTTAGAAATACATACAAACATGCTCTGCTGCAATTCGCAATTCGCAATTCGCTGCTAT 957
Qy 308 AlaLeuAsnLeuLeuValGlnGlyAlaGlyGlnValValSerValAspCysAsnPro--- 326
Db 958 AGGCAAAAGTCTTC-----CAAGTAATGAATCTAAGAGAGATGGAATG 1002
Qy 327 ---AlaGlnSerAlaLeuLeuGluLeuLysLysValAlaIleGlnGlnLeuGluPheGlu 345
Db 1003 TATGGGAAAGAAAGTTACTTGGAACTAATAGGACACGCTGTGGCTCATCTCTGAAGAAAGCT 1062
Qy 346 Asp-----ValTrpGlnLeuPheGly----- 352
Db 1063 GATGAGGCCAATGATAATCTCTCCGTGCTGTGTCTTCATTTGCGACGCTCCATGCTCTA 1122
Qy 353 -----GluGlyValHisProArgIleGluGluLeuTyrGluLysLysLeuAla 368
Db 1123 GCAGATCAGTATGAAGAGCAGAGTATTACTTCCAAAGGAATTCAGTAAGAGCTTACT 1182
Qy 369 ProPheLeuSerGlnThrSerHis-----AsnPheTrp----- 379
Db 1183 CCTAGCGAACAACACTGCTCCACTCTCGGTGATGGCAACTTTTCAGTGTACCAATGAAG 1242
Qy 380 -----SerLysArgLeuTyrPheGlnHisGlyLeu 390
Db 1243 TGTGAAGACAGGCGCCACCACTTTATAGAGGGTGTAT 1281
```

RESULT 13

US-10-932-182A-1448

```
/ Sequence 1448, Application US/10932182A
/ Publication No. US20060046253A1
/ GENERAL INFORMATION:
/ APPLICANT: NAKAO, YOSHIIRO
/ APPLICANT: NAKAMURA, NORIHISA
/ APPLICANT: KODAMA, YUKIKO
/ APPLICANT: FUJIMURA, TOMOKO
/ APPLICANT: ASHIKARI, TOSHIHIKO
/ TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
/ FILE REFERENCE: 030685-043
/ CURRENT APPLICATION NUMBER: US/10/932,182A
/ CURRENT FILING DATE: 2004-09-02
/ NUMBER OF SEQ ID NOS: 197023
/ SOFTWARE: PatentIn version 3.3
/ SEQ ID NO 1448
/ LENGTH: 7437
/ TYPE: DNA
/ ORGANISM: Saccharomyces pastorianus
US-10-932-182A-1448

Alignment Scores:
Pred. No.: 35.9 Length: 7437
Score: 101.50 Matches: 102
Percent Similarity: 32.0% Conservative: 71
Best Local Similarity: 18.9% Mismatches: 147
Query Match: 2.9% Indels: 221
DB: 7 Gaps: 27

US-10-620-914-45 (1-648) x US-10-932-182A-1448 (1-7437)
Qy 32 ThrValLeuArgHisMetTrpPheGlySer----- 41
Db 3670 TCAGTTTTGAAGAGCGCTTGGAAATCCAGCAACAAAGAACTAAAGAGGACTGGCAAGAA 3729
Qy 42 -----LysLysGlyAspAspHisAlaAlaArg--- 50
Db 3730 TGGAGTAAGCAGATATCTATTCATTCGTTAAAGAAATCCCTTCCCAAGCAGTGGAGCT 3789
Qy 51 -----LeuGluSerPheTyrGlyProGlnAla-----AlaAlaPhe 62
Db 3790 TGTTCAATCTCCTAGTAGTACTATTCCTAGTAGAGAACTCTTCAATCTGCTTTC 3849
Qy 63 AlaAlaArgLeuAlaGluArgSerAsnLeuIleTrpValAspLeuGlyGlyThrGly 82
Db 3850 GCAAGT-----ATTGGAGCGAGCTTTATAGCCAGTATCAA 3885
Qy 83 GluAsnVal-----AspMet 87
Db 3886 GAAGATTTGATCGAGTCGCTATGATGATGCTCTCTCTCAAAATCCGCGGAATA 3945
Qy 88 MetAlaAspTyrIleAspLeuAlaLysPhe-----LysSerIleTyrVal 102
Db 3946 CATCAAAACGTTATTAATTTGGTGAATTCATGAGCATGATGATAAGCATTTGCCTATC 4005
Qy 103 -----ValAspLeuCysHisSerLeuCysGluValAlaLys 114
Db 4006 CCAACTCAAAATTTGGTGAATATGCTGAGAGATGCCACCATAT----- 4050
Qy 115 LysLysAlaLysAlaLysGlyTrpLysAsnValGlnValValGluAlaAspAlaCysGln 134
Db 4051 -----GCTAAAGCATTTACATTACAAAGAAATTTAAATTTATCAA----- 4089
Qy 135 PheAlaProProGluGlyThrAlaThrLeuIlePheSerTyrSerLeuThrMetIle 154
Db 4090 -----GAGCTGAAAACCTCC-----ACATTTGATCACTACT 4119
Qy 155 ProProPheHisAsnValIleAspGlnAlaCysSerTyrLeuSerGlnAspGlyLeuVal 174
Db 4120 ATCAGTATCAATAATCAATTTGAACCAACG-----GACGCTGCAATT 4161
Qy 175 GlyValAlaAspPheTyrValSerGlyLysTyrAspLeuProLeuArgGlnMetProTrp 194
Db 4162 GGTATTTTAAAG---CACGCCAGCAACACCATTCGTTGCAATTTAAAGAAACATGCTTC 4218
```

QY 195 SerArgArgPhePheTrpArgSerIlePheAspIleAspAsnIleAspIleGlyProGlu 214
DB 4219 GAAAGAACTAGAAAGTGGGAGATGCACCTA----- 4248
QY 215 ArgArgAlaTyrLeuGluGlnLysLeuGluArgValTrpGluGlnAsnThrGlnGlySer 234
DB 4249 ---CACGCTTATTAATGAAGT-----GAGAAGCAGGAGATACATCT 4287
QY 235 IleProTyrValProTrpLeuArgAlaProTyrTyrValTrpIleGlyArgLeuProSer 254
DB 4288 GTAGGT-----GTTACTCTTGTAAGATGAGATCT 4317
QY 255 ValGlyHisAlaLeuHisGlu-----GluArgValGluArg 266
DB 4318 CTT---CATGCCCTAGTGAATGGGAGCAATTAATCTCAACTAGCTGCTAGAAATGGAGA 4374
QY 267 ProProMetPheProProThrPheLeuTyrThrGln----- 278
DB 4375 ATATCTAAATTCACGACGAGAAATTAATTTGCTCTTTGGCAGCGGTGCAGATGGGT 4434
QY 279 -----SerTrpGluAspProGluProAspMetGluValMetGluIleAsnProLys--- 295
DB 4435 TTAGGCGAATGGATATGCTTGGAGCAATATATTAGTGTATGAAACCAATCCCCAGAT 4494
QY 296 -----AspThrValLeuThrLeuThrSerGlyGly----- 305
DB 4495 AAGGAATCTCTTGATGCAATATATATTGTCACAAAAACGATPACGAGGATGCTGTATAA 4554
QY 306 -----CysAsnAlaLeuAsnLeuValGlnGlnGlyAlaGlyGlnValValSerValasp 323
DB 4555 CACATACTAAATGCCAGAGATTTGCTGGTAACCTGAAATATCCGCACCTAATTAACGAAAGT 4614
QY 324 CysAsnProAlaGlnSerAlaLeuLeuGluLeuLysValAlaIleGlnGlnLeuGlu 343
DB 4615 TACATAGAGCTTATAGTGAATTTGTAGAACACAAATATATCA-----GAG 4662
QY 344 PheGluAspValTrpGlnLeuPheGlyGlyGluValHisProArgIleGluGluLeuTyr 363
DB 4663 TTTGAAGAAATA-----ATTAAA 4680
QY 364 GluLysLeuAlaProPheLeuSerGlnThrSerHis-----AsnPheTrpSerLys 381
DB 4681 TATAAGCACTGGCGCTTAACCTCGGAGAAATACTTCAATATCAAAACCTTTGGACGAAA 4740
QY 382 ArgLeuTrpTyrPheGlnHisGlyLeuTyrTyrGlnGlyMetGlyLysLeuCysTrp 401
DB 4741 AGGTTG----- 4746
QY 402 ValLeuGlnCysLeuAlaValValLeuGlyLeuGlyLysThrValLysArgLeuAlaAsn 421
DB 4747 -----CTGGGTGGCAAAAGAACGTTGAT----- 4770
QY 422 AlaProThrMetGluGlnArgArgLeuTrpAspSerAsnMetLeuIleHisPheVal 441
DB 4771 -----TTATGGCAAAAGATGATCTCAGATGAGATCATTA 4803
QY 442 LysAsnGlyProLysProLeuValTrpLeuPheValLysPheValSerLeuVal----- 459
DB 4804 GTGATTAAACCAAAACAGATTACAAATATGGATCAAGTTTGCTTAATCTGTGTAGAAAA 4863
QY 460 -----LeuPheAsnLysAla-----ValLeuTrpPheGlyGlyVal 472
DB 4864 TCCGGTAGAATGAGACTGGCTCAAAAGGCACTGAATATGCTTCTTCAAGGAGCGGTGAT 4923
QY 473 Pro 473
DB 4924 CCT 4926

RESULT 14

US-10-932-182A-1448

; Sequence 1448, Application US/10932182A

; Publication No. US20060046253A1

GENERAL INFORMATION:
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHIRO
; APPLICANT: KODAMA, YUKIO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: 030685-043
; CURRENT APPLICATION NUMBER: US/10/932,182A
; CURRENT FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1448
; LENGTH: 7437
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-10-932-182A-1448

Alignment Scores:
Pred. No.: 35.9 Length: 7437
Score: 101.50 Matches: 102
Percent Similarity: 32.0% Conservative: 71
Best Local Similarity: 18.9% Mismatches: 147
Query Match: 2.9% Indels: 221
DB: 7 Gaps: 27

US-10-620-914-45 (1-648) x US-10-932-182A-1448 (1-7437)

QY 32 ThrValLeuArgHisMetTrpPheGlySer----- 41
DB 3670 TCAGTTTTGAAGAGCGCTTGGAAATTCACGCCAACAAAGAACTAAAGAGGACTGGCAAGAA 3729
QY 42 -----LysLysGlyAspAspHisAlaAlaArg--- 50
DB 3730 TGGAGTAAGCGATTATCTATTCAATTGCTTAAGAATCCCTTCCACGCACTGAGAGCT 3789
QY 51 -----LeuGluSerPheTyrGlyProGlnAla-----AlaAlaPhe 62
DB 3790 TGTTCGAATCTCGTAGTATGTAATCTCGTTAGTAGAAGAACTTTCAATCTGCTTTC 3849
QY 63 AlaAlaArgLeuAlaGluArgSerAsnLeuIleTrpValAspLeuGlyGlyGlyThrGly 82
DB 3850 GCAAGT-----ATTGGACGAGCTTTATAGCCAGTATCAA 3885
QY 83 GluAsnVal-----AspMet 87
DB 3886 GAAGATTGATCGAGTCGCTATGCATAGCTTCTCTCTCAAAATCCGCCGGAATA 3945
QY 88 MetAlaAspTyrIleAspLeuAlaLysPhe-----LysSerIleTyrVal 102
DB 3946 CATCAAAACGTTTATAAATTTGGTTCGAATTCATGGAGCATGATGAAGCAATTCCTATC 4005
QY 103 -----ValAspLeuCysHisSerLeuCysGluValAlaLys 114
DB 4006 CCAACTCAAAATTTGGGTGAATATGCTGAGAGATGCCGCAATAT----- 4050
QY 115 LysLysAlaLysAlaLysGlyTrpLysAsnValGlnValGluAlaAspAlaCysGln 134
DB 4051 -----GCTAAGCATTAATCAACAGAAATTAATTTATCAAA----- 4089
QY 135 PheAlaProGluGlyThrAlaThrLeuIleThrPheSerTyrSerLeuThrMetIle 154
DB 4090 -----GAGCCTGAAAACCTCC-----ACAATTTGAATCACTT 4119
QY 155 ProProPheHisAsnValIleAspGlnAlaCysSerTyrLeuSerGlnAspGlyLeuVal 174
DB 4120 ATCAGTATCAATAATCAATTTGAACCAACG-----GACGCTGCAATT 4161
QY 175 GlyValAlaAspPheTyrValSerGlyLysTyrAspLeuProLeuArgGlnMetProTrp 194
DB 4162 GGTATTTTAAAG---CACGCCAGCAGACACCATTCGTTGCAATTAAGGAAACATGGTTC 4218
QY 195 SerArgArgPheTrpArgSerIlePheAspIleAspIleGlyProGlu 214

Db 4219 GAAAACTAGAAAGTGGGAAGATGCACTA----- 4248
Qy 215 ArgArgAlaTyrLeuGluGlnLysLeuGluArgValTrpGluGlnAsnThrGlnGlySer 234
Db 4249 ---CACGCTTATAATGAACGT-----GAGAAGCGAGGAGATACATCT 4287
Qy 235 IleProTyrValProTyrLeuArgAlaProTyrTyrValTrpIleGlyArgLeuProSer 254
Db 4288 GTAGGT-----GTTACTCTTGTAAGATGAGATCT 4317
Qy 255 ValGlyHisAlaLeuHisGlu-----GluArgValGluArg 266
Db 4318 CTT---CATGCCCTAGTGAATGGAGCAATTATCTCAACTAGCTGCTAGAAAATGGAGA 4374
Qy 267 ProProMetPheProProThrPheLeuTyrThrGln----- 278
Db 4375 ATATCTAAATTTACCGACGAAAGAAATTAATTTCTCTTGGCAGCGGTGCAGCATGGGT 4434
Qy 279 -----SerTrpGluAspProGluProAspMetGluValMetGluIleAsnProLys--- 295
Db 4435 TTAGCGAATGGATATGCTTGGAGCAATATATTAGTGAATGAACCAAAATTTCCCGAGAT 4494
Qy 296 -----AspThrValLeuThrLeuThrSerGlyGly----- 305
Db 4495 AAGGAATTTCTTGTGATCAATATATTATTTGCACAAAAACGATTACGAGGATGCTGGTAAA 4554
Qy 306 -----CysAsnAlaLeuAsnLeuLeuValGlnGlyAlaGlyGlnValValSerValAsp 323
Db 4555 CACATACTAAATCCGAGAGATTTGCTGGTAATATATGCTGCAATATCCGCACTAATTAACGAAAGT 4614
Qy 324 CysAsnProAlaGlnSerAlaLeuLeuGluLeuLysLysValAlaIleGlnLeuGlu 343
Db 4615 TACAATAGACTTATAGTGAATTTGTTAGAACACAAATAATCACA-----GAG 4662
Qy 344 PheGluAspValTrpGlnLeuPheGlyGluGlyValHisProArgIleGluGluLeuTyr 363
Db 4663 TTGAAGAAATA-----ATTAAA 4680
Qy 364 GluLysLysLeuAlaProPheLeuSerGlnThrSerHis-----AsnPheTrpSerLys 381
Db 4681 TATAAGCAACTGGCGCTAACTCGGAGAAAAAACTTCATTTATCAAAACCTTTGGACGAAA 4740
Qy 382 ArgLeuTrpTyrPheGlnHisGlyLeuTyrTyrGlnGlyMetGlyLysLeuCysTrp 401
Db 4741 AGGTTC----- 4746
Qy 402 ValLeuGlnCysLeuAlaValValLeuGlyLysThrValLysArgLeuAlaAsn 421
Db 4747 -----CTGGTTGCCAAAGAACGTTGAT----- 4770
Qy 422 AlaProThrMetGluGluGlnArgArgLeuTrpAspSerAsnMetLeuIleHisPheVal 441
Db 4771 -----TTATGCCAAAGAGTACTCAGAGTGAGATCATTA 4803
Qy 442 LysAsnGlyProLysProLeuValTrpLeuPheValLysPheValSerLeuVal----- 459
Db 4804 GTGATTAACCAAAACAAGATTACAAATATGATCAAGTTTGTCTAATCTGTAGAAAA 4863
Qy 460 -----LeuPheAsnLysAla-----ValLeuTrpPheGlyGlyVal 472
Db 4864 TCCGGTAGAATGAGACTGGCTCAAAAGGCACTGAATATGCTTCTTGAAGGAGCGGTGAT 4923
Qy 473 Pro 473
Db 4924 CCT 4926

RESULT 15
US-11-169-041-29
; Sequence 29, Application US/11169041
; Publication No. US20060019284A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company

; TITLE OF INVENTION: IDENTIFICATION OF POLYNUCLEOTIDES FOR PREDICTING ACTIVITY OF
; TITLE OF INVENTION: COMPOUNDS THAT INTERACT WITH AND/OR MODULATE PROTEIN TYROSINE
; TITLE OF INVENTION: KINASES AND/OR PROTEIN TYROSINE KINASE PATHWAYS IN LUNG CANCER
; TITLE OF INVENTION: CELLS
; FILE REFERENCE: 10001 NP
; CURRENT APPLICATION NUMBER: US/11/169,041
; PRIOR FILING DATE: 2005-06-28
; PRIOR APPLICATION NUMBER: 60/584,405
; PRIOR FILING DATE: 2004-06-30
; NUMBER OF SEQ ID NOS: 527
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 29
; LENGTH: 3063
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-169-041-29

Alignment Scores: 12.5 Length: 3063
Pred. No.: 101.00 Matches: 90
Score: 33.3% Conservative: 54
Percent Similarity: 20.8% Mismatches: 163
Best Local Similarity: 2.9% Indels: 126
Query Match: 12 Gaps: 19
DB: 19

US-10-620-914-45 (1-648) x US-11-169-041-29 (1-3063)

Qy 12 TyrThrLysLysAsnPheSerLeuGluLysLeuLysSerSerMetLysAspAspLeu 31
Db 198 TTTTACCGAGCTGAGTGTTCAGATCCGTAATCTCAAA---GCCACAATGCAACCTACTG 254
Qy 32 ThrValLeuArgHisMetTrpPheGlySerLysLysGlyAspAspHisAlaAla----- 49
Db 255 GCCTATCTTAAGCACCTC-----AAAGGCAAAACGAGGCGAGCTGGAA 299
Qy 50 -----ArgLeuGluSerPheTyrGlyProGlnAlaAla-----AlaPheAlaAla 64
Db 300 TGCTTACGTAAGCTGAAGAGTTAATCCAGCAAGAGCATGCTGACCAGGCGAGAAATCAGA 359
Qy 65 ArgLeuAlaGluArgSerHisLeuIleTrpValAspLeuGlyGlyGlyThrGlyGluAsn 84
Db 360 AGTCTGTCACCTGGGGAACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 419
Qy 85 ValAspMetMetAlaAspTyrIleAsp-----LeuAlaLysPheLysSerIleTyrVal 102
Db 420 GTTCAGATTATGTAGCAAGGTGAGACATGCTGTGAGAGATTTTCCAGTCCCTATAGA 479
Qy 103 ValAspLeuCysHisSerLeuGluValAlaLysLysLysAlaLysLysGlyTrp 122
Db 480 ATTGAGAGTCCAGAGCTTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 539
Qy 123 LysAsn-----ValGlnValValGluAlaAspAlaCysGlnPheAlaPro-----Pro 138
Db 540 CAAAATGAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 599
Qy 139 GluGlyThrAlaThrLeuIleThrPheSerTyrSerLeuThrMetIleProProPheHis 158
Db 600 GAATTCACCTCTGAGCTGGCAATAGCAAGTACCGTCTGGACCACTGGCCACCATCTCAG 659
Qy 159 AsnValIleAspGlnAlaCysSerTyrLeuSerGlnAspGlyLeuValGlyValAlaAsp 178
Db 660 AAGCCATTGAC-----CCTCTGAGGCAAGCC----- 686
Qy 179 PheTyrValSerGlyLysTyrAspLeuProLeuArgGlnMetProTrpSerArgArgPhe 198
Db 672 -----CCTCTGAGGCAAGCC----- 686
Qy 199 PheTrpArgSerIlePheAspIleAspAsnIleAspIleGlyProGluArg----- 215
Db 687 -----ATTCCGCTGAATCTCTGACACCACTGACCTT 716
Qy 216 ArgAlaTyrLeuGluGlnLysLeuGluArgValTrpGluGlnAsnThrGlnGlySerIle 235

```
Db 717 AAGTCTCTCTGGCTCTGAAGCTTCATAGATCGGTGAAGAA----- 758
QY 236 ProTyrValProTyrLeuArgAlaProTyrTyrValTyrIleGlyArgLeuProSerVal 255
Db 759 -----GGTGAAGAGGAAGGTGAA 776
QY 256 GlyHisAlaLeuHisGluGluArgValGluArgProPro-----MetPhePro 271
Db 777 GGAGAGAAGTTAGTTGAAGAAGCTTTGGAGAAAGCCCGAGGTGTAACAGATGTACTTCGC 836
QY 272 ProThrPheLeuTyrThrGlnSerTyrGlnSerTyrGluAspProAspMetGluVal----- 289
Db 837 AGTGCAGCCCAAGTTTATCGAAGAAAGATGACGACAGCAAGCGATTGAACGTCTTAAA 896
QY 290 -----MetGluIleAsnProLysAspThrValLeuThrLeuThrSerGlyGlyCysAsn 307
Db 897 AAGGCTTTAGAAATACATACCAACAATGCCCTACTGCAATGCCCAATTTGGTGCTGTAT 956
QY 308 AlaLeuAsnLeuValGlnGlyAlaGlyGlnValValSerValAspCysAsnPro--- 326
Db 957 AGGGCAAAAGCTCTTC-----CAAGTAATGAATCTAAGAGAGAATGGAATG 1001
QY 327 ---AlaGlnSerAlaLeuLeuGluLeuLysLysValAlaIleGlnGlnLeuGluPheGlu 345
Db 1002 TATGGGAAAGAAAGTTACTTGGAACTTAATAGGACACGCTGTGGCTCATCTGAAGAAAGCT 1061
QY 346 Asp-----ValTyrGlnLeuPheGly----- 352
Db 1062 GATGAGGCCAATGATATCTCTCCGTGTCTGTTCATCTTCCAGCCTCCATGCTCTA 1121
QY 353 -----GluGlyValHisProArgIleGluGluLeuTyrGluLysLysLeuAla 368
Db 1122 GCAGATCAGTATGAAGAAGCAGAGTATTACTTCCAAAAGGAATTCAGTAAAGAGCTTACT 1181
QY 369 ProPheLeuSerGlnThrSerHis-----AsnPheTrp----- 379
Db 1182 CCGTAGCGGAACAACACTGCTCCATCTGCGGTATGGCAACTTTCAGCTGTACCAATGAAG 1241
QY 380 -----SerLysArgLeuTyrPheGlnHisGlyLeu 390
Db 1242 TGTGAAGACAGGCCCATCCACCACTTTATAGAGGGTGTA 1280
```

Search completed: March 14, 2006, 07:27:55
Job time : 1251 secs

THIS PAGE BLANK (USPTO)